

THE ANNALS
of
MATHEMATICAL
STATISTICS

(FOUNDED BY H. C. CARYER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF
MATHEMATICAL STATISTICS

VOLUME XV

1944

THE ANNALS OF MATHEMATICAL STATISTICS

EDITED BY
S. S. WILKS, *Editor*

A. T. CRAIG
W. E. DEMING
T. C. FRY

H. HOTELLING
J. NEYMAN
W. A. SHEWHART

WITH THE COÖPERATION OF

W. G. COCHRAN
J. H. CURTIS
H. F. DODGE
P. S. DWYER

C. EISENHART
W. K. FELLER
P. G. HOEL
W. G. MADDOW

A. M. MOOD
H. SCHEFFÉ
A. WALD
J. WOLFOWITZ

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore 2, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore 2, Md., or to the Secretary of the Institute of Mathematical Statistics, P. S. Dwyer, 116 Rackham Hall, University of Michigan, Ann Arbor, Mich.

Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December. Because of war-time difficulties of publication, issues may often be from two to four weeks late in appearing. *Subscribers are therefore requested to wait at least 30 days after month of issue before making inquiries concerning non-delivery.*

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is \$5.00 per year. Single copies \$1.50. Back numbers are available at \$5.00 per volume, or \$1.50 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

CONTENTS OF VOLUME XV

ARTICLES

ANDERSON, T. W. AND GIRSHICK, M. A. Some Extensions of the Wishart Distribution	345
BANCROFT, T. A. On Biases in Estimation Due to the Use of Preliminary Tests of Significance	190
CHURCHMAN, C. WEST AND EPSTEIN, BENJAMIN. On the Statistics of Sensitivity Data	90
DIXON, WILFRID J. Further Contributions to the Problem of Serial Correlation	119
DOOB, J. L. The Elementary Gaussian Processes	229
DWYER, PAUL S. A Matrix Presentation of Least Squares and Correlation Theory with Matrix Justification of Improved Methods of Solution	82
EPSTEIN, BENJAMIN AND CHURCHMAN C. WEST. On the Statistics of Sensitivity Data	90
GEIRINGER, HILDA. On the Probability Theory of Linkage in Mendelian Heredity	25
GIRSHICK, M. A. AND ANDERSON, T. W. Some Extensions of the Wishart Distribution	345
GRUBBS, FRANK E. On the Distribution of the Radial Standard Deviation	75
GUMBEL, E. J. Ranges and Midranges	414
HALMOS, PAUL R. Random Alms	182
HARSHBARGER, BOYD. On the Analysis of a Certain Six-by-Six Four-Group Lattice Design	307
HOTELLING, HAROLD. Some Improvements in Weighing and other Experimental Techniques	207
Hsu, L. C. Some Combinatorial Formulas with Applications to Probable Values of a Polynomial-Product and to Differences of Zero	399
HURWITZ, H. AND KAC, M. Statistical Analysis of Certain Types of Random Functions	173
KAC, M. AND HURWITZ, H. Statistical Analysis of Certain Types of Random Functions	173
LEIMER, EMMA. Inverse Tables of Probabilities of Errors of the Second Kind	388
LEVENE, H. AND WOLFOWITZ, J. The Covariance Matrix of Runs Up and Down	58
MADOW, LILLIAN AND MADOW, WILLIAM G. On the Theory of Systematic Sampling, I.	1
MADOW, WILLIAM G. AND MADOW, LILLIAN. On the Theory of Systematic Sampling, I.	1
ROBBINS, H. E. On the Measure of a Random Set	70

SATTERTHWAITE, F. E. Error Control in Matrix Calculation	373
ULLMAN, JOSEPH. The Probability of Convergence of an Iterative Process of Inverting a Matrix	205
WALD, ABRAHAM. On a Statistical Problem Arising in the Classification of an Individual into One of Two Groups	145
WALD, ABRAHAM. On Cumulative Sums of Random Variables	283
WALD, A. AND WOLFOWITZ, J. Statistical Tests Based on Permutations of the Observations	358
WOLFOWITZ, J. AND LEVENE, H. The Covariance Matrix of Runs Up and Down	58
WOLFOWITZ, J. Asymptotic Distributions of Runs Up and Down	163
WOLFOWITZ, J. AND WALD, A. Statistical Tests Based on Permutations of the Observations	358

NOTES

BIRNBAUM, J. W. AND ZUCKERMAN, HERBERT S. An Inequality Due to H. Hornich	328
BOWKER, ALBERT H. Note on Consistency of a Proposed Test for the Problem of Two Samples	98
COURT, LEWIS M. A Reciprocity Principle for the Neyman-Pearson Theory of Testing Statistical Hypotheses	326
GREVILLE, T. N. E. A Generalization of Waring's Formula	218
GREVILLE, T. N. E. On Multiple Matching with One Variable Deck	432
HOTELLING, HAROLD. Note on a Matrix Theorem of A. T. Craig	427
KIMBALL, BRADFORD F. Note on Asymptotic Value of Probability Distribution of Sum of Random Variables which are Greater than a Set of Arbitrarily Chosen Numbers	423
LANDAU, H. G. Note on the Variance and Best Estimates	219
LONSETH, A. T. On Relative Errors in Systems of Linear Equations	323
ROBBINS, HERBERT. On Distribution-free Tolerance Limits in Random Sampling	214
ROBBINS, H. E. On the Expected Values of Two Statistics	321
SCHIEFFE, H. AND TUKEY, J. W. A Formula for Sample Sizes for Population Tolerance Limits	217
SCHIEFFE, HENRY. A Note on the Behrens-Fisher Problem	430
TUKEY, J. W. AND SCHIEFFE, H. A Formula for Sample Sizes for Population Tolerance Limits	217
WALD, ABRAHAM. Note on a Lemma	330
WILKINS, J. ERNEST, JR. A Note on Skewness and Kurtosis	333
WOLFOWITZ, J. Note on Runs of Consecutive Elements	97
ZUCKERMAN, HERBERT S. AND BIRNBAUM, J. W. An Inequality due to H. Hornich	328

MISCELLANEOUS

Abstracts of Papers.	342
Amendments to the Constitution and By-Laws of the Institute	340
Annual Report of the President of the Institute	111
Annual Report of the Secretary-Treasurer of the Institute	113
Constitution and By-Laws of the Institute	111
Directory of the Institute	411
Henry Lewis Rietz--In Memoriam. By A. R. Crathorne	102
News and Notices	109, 222, 336, 439
Report on the Washington Meeting of the Institute	226
Report on the Wellesley Meeting of the Institute	338
Review of "Student's" Collected Papers W. G. Cochran.	435

ON THE THEORY OF SYSTEMATIC SAMPLING, I

BY WILLIAM G. MADOW AND LILLIAN H. MADOW^{1,2}

1. Introduction. It is no longer necessary to demonstrate a need for the theory of designing samples. Many of the policy and operating decisions of both government and private industry are based on samples. There has been an increasing tendency in government and industry to make use of sampling theory.³

Unfortunately there are still considerable differences between the theory and practice of sampling. The origins of these differences are, on the one hand, the ignorance of administrators concerning the practical contributions that sampling theory can make, and on the other, the lack of sampling theory permitting the evaluation of some useful sampling designs.

Much has been and is being done towards bringing theory and practice into agreement.⁴ Administrators and samplers are each successfully educating the others. However, there still exist sampling designs for which an adequate theory has not been developed, even though experience indicates that if such a theory were developed it would demonstrate the superiority of those designs over others for which a theory has been developed.

Perhaps the major omission of sampling theory today is the lack of any statistical method for reaching a decision on whether to take a completely random sample of n elements of a population of N elements, or to take a systematic sample, that is, to begin with element i , and select elements $i, i + k, \dots, i + (n - 1)k$, as the sample, the starting point i being chosen at random and $N = kn$ approximately.⁵ It is with respect to this question of whether to take a systematic⁶ or random sample that the statistician is in a dilemma because he has the alternative of recommending a systematic sampling procedure for which no theory exists, or a random sampling procedure that may well yield worse

¹ Bureau of Agricultural Economics and Food Distribution Administration, U. S. Department of Agriculture, Washington, D. C.

² Presented at a meeting of the seminar in statistics of the Graduate School, U. S. Department of Agriculture, November 2, 1943.

³ The recognition of the need for statisticians who know sampling theory has resulted in courses in sampling being given in some of the colleges and universities.

⁴ One need only refer to the recent development of positions, the duties of which include giving advice on sampling techniques as well as working in the field of application.

⁵ In this paper we will assume that $N = kn$. To do away with that assumption would not add much in the way of generality while it would require some fairly detailed discussion. It may be remarked that when N is not exactly kn , then systematic sampling procedures in which all starting points have equal probability of selection are biased, although the bias is usually trivial. If N is known this bias can be removed by sampling proportionate to possible size of systematic sample.

⁶ As we define systematic sampling procedures, a systematic sampling procedure is a random sampling procedure in which many of the C_N^n selections of n from N items are excluded.

results than the systematic procedure. The purpose of this paper is to resolve that conflict by providing an adequate theory of systematic sampling.

In the following sections we present the first parts of our research in the theory of systematic samples. Although this research covers both the theory of sampling single elements and sampling clusters of elements, we shall consider, in this paper, the sampling units to be single elements, not clusters of elements. The latter problem will be dealt with in a later paper. We shall present the theory of systematic sampling both from an unstratified population and a stratified population. Formulas for the mean value and variances of the estimates are derived. Comparisons with random and stratified random sampling designs are made. Furthermore, the estimates of the variances and formulas for "optimum" size and allocation of samples are derived.

A fundamental part of the analysis is the demonstration that from a knowledge of the variance of the population⁷ and certain serial correlations or serial variances, can be estimated the variance of estimates based on systematic samples. The basic results are:

- a. if the serial correlations have a positive sum, systematic sampling is worse than random sampling,
- b. if the serial correlations have a sum that is approximately zero, systematic sampling is approximately equivalent to random sampling, and
- c. if the serial correlations have a negative sum, systematic sampling is better than random sampling.

2. The use of a finite population. In this paper we assume, for the calculation of the expected values, that we are sampling from a finite population of elements even though the size of the population may be large enough to permit the use of limiting distributions. Often, this is, mathematically, a matter of choice. The same results would be obtained by assuming a correctly defined multivariate normal distribution and using the notions of conditional probability. From a physical point of view, however, there are several factors that lead to the use of the finite population. We are most frequently sampling an existing population whose laws of transformation are either unknown or not mathematically expressed.⁸ Consequently, the notion of a normal or other specified distribution from which we sample and use conditional probability is not part of our thinking concerning the physical problem. On the other hand, if we consider the population to be a finite population, and use a table of random numbers to draw our sample from the finite population, we are using only mathematics implicit in our physical problem. Furthermore, we do obtain a repeatable experiment, that of selecting a random number, that we know is in a state of statistical control.

In the usual problems of the theory of random sampling, the number of

⁷ By "variance of population" without further qualification is meant the variance of a random sample of one element of the population.

⁸ In other words, our population is not in a state of statistical control over time.

possible samples yielding different sample means is large enough so that the sample means may, with a sufficiently large size of population and sample, be expected to be approximately normally distributed. In systematic sampling, however, the number of possible sample means is usually very small and even if the sizes of population and sample are large, it is difficult to assume a normal distribution. Consequently, in our interpretation of the means and variances of systematic samples we are led to regard the elements of our populations as being the results of single observations on random variables, the distributions of which may vary from element to element. The interpretations that we then make become interpretations of conditional probability, and if the sizes of population and sample are sufficiently large, we can assume that the arithmetic mean of each of the possible sample means is normally distributed.

The theory of systematic sampling under the assumption of an appropriate normal multivariate distribution will be dealt with at a later time.

3. Definitions. Let the finite population to be sampled consist of N elements, x_1, \dots, x_N .

By a sample design is meant the combination of a method of classifying these N elements into k classes that may or may not overlap, and a method of selecting one of these k classes, each class having a designated probability of being selected. The sampling procedure associated with a given sample design is the operation of selecting one of the k classes according to the method stated in the sample design. The sample is the particular class obtained by the sampling procedure.

By a random sampling procedure is meant any sampling procedure such that if the sampling design yields k classes then the probability of selecting anyone of these classes is $1/k$. Any sample design having a random sampling procedure associated with it is a random sampling design. One of the nonrandom sampling procedures that is being used is the procedure in which the classes have associated to them numbers, called sizes, and the probability of a given class being the sample is proportionate to its size.⁹ Other nonrandom sampling procedures are doubtless being used.

By an unrestricted random sampling design for selecting n elements from N elements is meant the sampling design such that there are C_n^N classes, the possible selections of n from N elements, each having a probability of $1/C_n^N$ of being the sample. The associated random sampling procedure might consist in identifying each class by a number i , $i = 1, \dots, C_n^N$ and selecting a number i from a table of random numbers. The random sampling procedure might also consist in identifying the N elements with numbers $j = 1, \dots, N$, and then selecting a number j from a table of random numbers, then selecting a different number j from a table of random numbers, and following that procedure until n numbers

⁹ For a discussion of this problem see the paper entitled, "On the theory of sampling from finite populations," by Morris H. Hansen and William N. Hurwitz, *Annals of Math. Stat.*, Vol. 14 (1943), pp. 333-362.

from $1, \dots, N$ without repetition have been selected from the table of random numbers. The elements associated with these integers would be a random sample. It is easy to see that the two procedures are equivalent.

A random sampling design that is not unrestricted is said to be restricted. There are many types of restricted random sampling designs of which what we call systematic designs are only one. Among these restricted designs are stratified, cluster, double, matched, polynomial, and other sampling designs, each having been developed as attempts to bring theory and practice together, to suggest improvements in practice, and to solve problems arising in practice.

By a systematic sampling design is meant a classification of the N elements into k classes, S_1, \dots, S_k where S_i consists of $x_i, x_{i+k}, \dots, x_{i+(n-1)k}$, and a random sampling procedure for selecting one of the S_i .

It is thus clear that a systematic sampling design is a type of cluster sampling design. It will be shown that the new aspect of cluster sampling introduced in systematic sampling is that a knowledge of the order of the elements in the population is used to obtain the values of the intraclass correlation coefficient and changes in the value of that coefficient as the size of sample changes.

Sampling designs may involve combinations of random and systematic sampling designs, as well as random and nonrandom sampling procedures.

The population from which these samples are drawn may or may not be stratified and the sampling units may be single elements or clusters of elements.

4. Bases for selecting among sample designs. From the many sampling designs that can be constructed in order to obtain desired estimates, one will be chosen for use on the bases of administrative considerations, cost, and sampling error. It has become customary, on the basis of limiting distribution theory and the theory of best linear unbiased estimates to use the standard deviation of the sample estimate about the character estimated as the measure of sampling error.

Although in this paper we shall continue this practice, it must be pointed out that as more sampling designs are constructed, there is the danger that for some of these designs the limiting distribution theory is not valid, and the use of the standard error becomes more a matter of custom than the result of analysis. This danger is present for systematic sampling designs and is being further investigated.

It is perhaps desirable to remark that bias, consistency, and efficiency are properties of the sampling design and estimation functions used, not of the particular sample obtained. Any estimate based on a sample will probably differ from the character estimated. It is the function of statistical analysis to indicate how large this difference may be.

5. Notation. The letter, P , with appropriate subscripts is used for population, and subpopulations such as strata.

The number of strata is denoted by L , and the number of elements in the i^{th} stratum is denoted by N_i . Sizes of sample are denoted by n with appropriate subscripts.

The arithmetic mean of the elements of a population or subpopulation is denoted by \bar{x} with appropriate subscripts.

Any particular subclass of a population as defined by the sampling design is denoted by S with subscripts. Estimates based on an S with subscripts are denoted by \bar{x} with subscripts.

6. Unstratified systematic sampling, the sampling unit consisting of one element. The values assumed by the subscripts used in this section are given in Appendix A

Let the population, P , consist of N elements x_1, \dots, x_N . It is desired to estimate the arithmetic mean, \bar{x} , of P .

Let $N = kn$, and let the class S_i consist of the n elements $x_i, x_{i+1}, \dots, x_{i+(n-1)k}$. Then; the systematic sampling design for estimating \bar{x} from a sample of size n , consists of the k classes, S_1, \dots, S_k , and the requirement that the sampling procedure be such that the probability is $1/k$, that S_i is the class selected by the sampling procedure.

Let \bar{x}_i be the arithmetic mean of the elements of S_i , i.e., $n\bar{x}_i = x_i + x_{i+k} + \dots + x_{i+(n-1)k}$, and let \bar{x} be the sample mean, i.e., $\bar{x} = \bar{x}_i$ if S_i is selected by the sampling procedure.

In dealing with systematic sampling, we shall have occasion to use both the circular and non-circular definitions of the serial correlation coefficients and the associated serial variances.

We shall assume that if $h > kn$ then $x_h = x_{h-kn}$. This is used in the circular definitions.

$$\text{Let } kn\sigma^2 = \sum_v (x_v - \bar{x})^2,$$

$$\text{and let } knC_{k\mu} = \sum_v (x_v - \bar{x})(x_{v+k\mu} - \bar{x}).$$

Then, the circular definition of the serial correlation coefficient $\rho_{k\mu}$ is $\sigma^2 \rho_{k\mu} = C_{k\mu}$, which we shall use unless n is even, when we define $\rho_{kn/2}$ by the equation

$$2\sigma^2 \rho_{kn/2} = C_{kn/2},$$

in order to simplify the writing of the formula for σ_x^2 .

Similarly, if we define the serial variance, $s_{k\mu}$, by the equation $kn s_{k\mu} = \sum_v (x_v - x_{v+k\mu})^2$, then we are using the circular definition of the serial variance. The circular definition of the serial variance ratio $v_{k\mu}$ is then $\sigma^2 v_{k\mu} = s_{k\mu}$ which we shall use unless n is even, when we define $v_{kn/2}$ by the equation

$$2\sigma^2 v_{kn/2} = s_{kn/2}.$$

The non-circular definitions of the serial correlations and serial variances are given by

$$(1) \quad \begin{aligned} k(n - \delta)C'_{ks} &= \sum_j (x_j - \bar{x})(x_{j+ks} - \bar{x}), \\ \sigma^2 \rho'_{ks} &= C'_{ks}, \\ k(n - \delta)s'_{ks} &= \sum_j (x_j - x_{j+ks})^2, \end{aligned}$$

and

$$\sigma^2 t'_{ks} = s'_{ks}.$$

The intraclass correlation coefficient $\bar{\rho}_k$ is defined by the equation

$$\sigma^2 \bar{\rho}_k = S(x_\mu - \bar{x})(x_\nu - \bar{x}),$$

where the random process consists in first sampling one of the S , at random and then selecting two of the x 's at random from the S , that was selected. Then, since

$$k\sigma_x^2 = \sum_i (\bar{x}_i - \bar{x})^2,$$

and,

$$(2) \quad \sigma^2 \bar{\rho}_k = (n/n - 1)\sigma_x^2 - (1/n - 1)\sigma^2$$

we have

$$(3) \quad \sigma_x^2 = \frac{1}{n} (1 + (n - 1)\bar{\rho}_k)$$

It is easy to see from (1) that the intraclass correlation coefficient is given by

$$\begin{aligned} \bar{\rho}_k &= \frac{2}{n(n - 1)} \sum_s (n - \delta) \rho'_{ks} \\ &= \frac{2}{n - 1} \sum_\mu \rho_{k\mu}, \end{aligned}$$

and that consequently, if n is odd, $\bar{\rho}_k$ is the arithmetic mean of the $\rho_{k\mu}$ while if n is even, $\bar{\rho}_k$ is equal to the arithmetic mean of the $\rho_{k\mu}$ multiplied by $n/(n - 1)$.

THEOREM¹⁰: Using the systematic sampling design, the estimate \bar{x} is an unbiased estimate of \bar{x} , and has variance $\sigma_{\bar{x}}^2$ where

$$\begin{aligned} (4) \quad \sigma_{\bar{x}}^2 &= \sigma^2 \left\{ 1 - \frac{1}{n^2} \sum_i (n - \delta) v'_{ki} \right\} \\ &= \sigma^2 \left(1 - \frac{1}{n} \sum_\mu v_{k\mu} \right) \\ &= \frac{\sigma^2}{n} \left\{ 1 + \frac{2}{n} \sum_s (n - \delta) \rho'_{ks} \right\} \\ &= \frac{\sigma^2}{n} (1 + 2 \sum_\mu \rho_{k\mu}) \\ &= \frac{\sigma^2}{n} \{ 1 + (n - 1)\bar{\rho}_k \}. \end{aligned}$$

¹⁰ A proof of Theorem 1 that is somewhat simpler to follow but which, in the authors opinion, is not as informative as that given below could be obtained by substituting for $\bar{\rho}_k$ using equations (2) and (3). The lemmas in Appendix B are, of course, of interest in themselves in finite sampling.

PROOF: From the definitions of expected value, \bar{x}_1 , \bar{x} , and the systematic sampling design, it follows that \bar{x} is a variate with possible values $\bar{x}_1, \dots, \bar{x}_k$, the probability that $\bar{x} = \bar{x}_i$ being $1/k$. Then

$$(5) \quad k\bar{x} = \bar{x}_1 + \dots + \bar{x}_k,$$

and, when the values of the \bar{x}_i are substituted in (5), it follows that $\bar{x} = \bar{x}$, that is, \bar{x} is an unbiased estimate of \bar{x} .

Having calculated \bar{x} , it is necessary to calculate \bar{x}^2 in order to evaluate $\sigma_{\bar{x}}^2$. From the definition of expected values, it follows that

$$(6) \quad k\bar{x}^2 = \bar{x}_1^2 + \dots + \bar{x}_k^2,$$

and when the values of the \bar{x}_i are substituted in (6), it follows that

$$(7) \quad n^2 k \bar{x}^2 = \sum_{i, \alpha, \gamma} x_{i+(\alpha-1)k} x_{i+(\gamma-1)k}$$

Then, when $f(u)$ is replaced by u in Lemma 6 of Appendix B, it follows, from the definition of the variance, that $\sigma_{\bar{x}}^2 = \left(\frac{1}{kn}\right) \sum_r (x_r - \bar{x})^2 - \left(\frac{1}{kn^2}\right) \sum_{i,j} (x_i - x_{j+k})^2 = \sigma^2 - \frac{1}{n^2} \sum_i (n - \delta) s'_{ki}$, and when $f(u)$ is replaced by u in Lemma 8, it follows that

$$\begin{aligned} \sigma_{\bar{x}}^2 &= \left(\frac{1}{kn^2}\right) \sum_r (x_r - \bar{x})^2 + \left(\frac{2}{kn^2}\right) \sum_{i,j} (x_i - \bar{x})(x_{j+k} - \bar{x}) \\ &= \left(\frac{1}{n}\right) \sigma^2 + \frac{2}{n^2} \sum_i (n - \delta) c'_{ki}. \end{aligned}$$

If in Lemma 9 of Appendix B we now replace $f(x_j, x_{j+k})$ by $(x_j - x_{j+k})^2$ then $\sigma_{\bar{x}}^2 = \sigma^2 - \left(\frac{1}{n}\right) \sum_{\mu} s_{k\mu}$,

and if we replace $f(x_j, x_{j+k})$ by $(x_j - \bar{x})(x_{j+k} - \bar{x})$ then

$$\sigma_{\bar{x}}^2 = \frac{1}{n} (\sigma^2 + 2 \sum_{\mu} c_{k\mu}).$$

Finally, we have, then

$$\sigma_{\bar{x}}^2 = \sigma^2 \left(1 - \frac{1}{n} \sum_{\mu} v_{k\mu}\right),$$

and

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} (1 + 2 \sum_{\mu} \rho_{k\mu}).$$

7. Possible values of the $\rho'_{k\delta}$, $\rho_{k\mu}$ and $\sigma_{\bar{x}}^2$. Let us investigate briefly the effects of different patterns of variation on the values of $\rho'_{k\delta}$ and $\sigma_{\bar{x}}^2$. Now $\sigma^2 \rho'_{k\delta} = \frac{1}{k(n-\delta)} \sum_j (x_j - \bar{x})(x_{j+k} - \bar{x})$. Suppose that $x_i = x_{i+k}$, $\delta = 1, \dots, n-1$,

$i = 1, \dots, k$. Then $\sum_i (x_i - \bar{x})^2 = n \sum_i (x_i - \bar{x})^2$, and $\sum_i (x_i - \bar{x})(x_{i+k} - \bar{x}) = (n - \delta) \sum_i (x_i - \bar{x})^2$. Upon substitution it follows that $\rho'_{ki} = 1$, and $\sigma^2_{\bar{x}} = \sigma^2$.

This result for $\sigma^2_{\bar{x}}$ is intuitively clear, since all the variability is among the possible samples, and thus any particular systematic sample is equivalent to one observation.

Suppose, on the other hand, that $x_{k\delta+\alpha} = x_{k\delta+\beta} \alpha, \beta = 1, \dots, k; \delta = 1, \dots, n - 1$. Then $\sum_i (x_i - \bar{x})^2 = k \sum_{\alpha} (x_{i+(\alpha-1)k} - \bar{x})^2$ for any $i, i = 1, \dots, k$, and $\sum_i (x_i - \bar{x})(x_{i+k} - \bar{x}) = k \sum_{\alpha} (x_{i+(\alpha-1)k} - \bar{x})(x_{i+(\alpha-1)k} - \bar{x})$. Furthermore $0 = [\sum_{\alpha} (x_{i+(\alpha-1)k} - \bar{x})]^2 = \sum_{\alpha} (x_{i+(\alpha-1)k} - \bar{x})^2 + 2 \sum_{\lambda, \delta} (x_{i+(\lambda-1)k} - \bar{x})(x_{i+(\lambda-1)k} - \bar{x})$. Hence

$$2 \sum_i \frac{n - \delta}{n} \rho'_{ki} = -1 \quad \text{and} \quad \sigma^2_{\bar{x}} = 0.$$

It is possible to construct examples in which any particular $\rho'_{ki} = -1$, but in such cases the remaining ρ'_{ki} each vanish. It is well known that the minimum value of $\bar{\rho}_k$ is $-1/(n - 1)$.

Finally, let us consider the expected values of ρ'_{ki} and $\sigma^2_{\bar{x}}$ if the x 's have been assigned their subscripts at random. These values are $\rho'_{ki} = -1/(nk - 1)$ and $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} \left(\frac{nk - n}{nk - 1} \right)$.

In most practical applications of systematic sampling it will be highly unlikely that the distribution of the x 's will be such that the x 's may be said to have been assigned their subscripts at random. In general, there will be logical reasons to expect that the x 's will have some fundamental trend. Thus, information will often be available, or may be obtained by a small subsample, on the basis of which a decision can be made to use some approach differing from that of assuming the subscripts of the x 's to have been assigned at random.

8. Estimates of the parameters. The formulae obtained in section 6 for the variance of the mean of a systematic sample are population formulae. Their values depend on the values of all the elements of the population. However, even in tests of possible sampling procedures, we rarely have available the resources with which to study the entire population. Consequently, it becomes necessary to investigate the possibility of estimating the population variances and serial correlations from samples. It will be shown that the estimates of the variances and correlations derived from a single S_i are biased and inconsistent whereas it will be possible to construct unbiased or consistent estimates from samples of more than one of the S_i . The sampling variations of these estimates must be left for further study.

Let us assume that instead of sampling only one of the S_i , as we did in section

6, we sampled g of the S_i at random. Then our sample would consist of all the elements in the S_β . The sample mean, \bar{x} , is defined by

$$g\bar{x} = \sum_{\beta} \bar{x}_{\beta}$$

if the subscripts of our sample classes are i_1, \dots, i_g .

Then it is easy to see that \bar{x} is unbiased. Furthermore, since we can regard this sampling procedure as the sampling of g of k elements at random, it follows that $\sigma_{\bar{x}}^2 = \frac{k-g}{k-1} \frac{1}{g} \sigma_x^2$ and, we have evaluated σ_x^2 in section 6.¹¹

Since

$$k\sigma_x^2 = \sum_i (\bar{x}_i - \bar{x})^2,$$

we shall consider estimating σ_x^2 by s_g^2 where

$$gs_g^2 = \sum_{\beta} (\bar{x}_{\beta} - \bar{x})^2.$$

Now, since $E \bar{x}^2 = \sigma_{\bar{x}}^2 + \bar{x}^2$, and

$$E \sum_{\beta} \bar{x}_{\beta}^2 = \frac{g}{k} \sum_i \bar{x}_i^2 = g(\sigma_{\bar{x}}^2 + \bar{x}^2)$$

it follows that $E s_g^2 = \sigma_x^2$, and hence s_g^2 is an unbiased estimate of σ_x^2 . Furthermore $E s_g^2 = \frac{g(k-1)}{k-g} \sigma_{\bar{x}}^2$.

We now turn to estimates of the $\rho_{k\mu}$ and σ^2 .

Let

$$gn\delta_g^2 = \sum_{\beta, \alpha} (x_{\beta+(\alpha-1)k} - \bar{x})^2$$

and let

$$gnc_{k\mu g} = \sum_{\beta, \alpha} (x_{\beta+(\alpha-1)k} - \bar{x})(x_{\beta+(\alpha+\mu-1)k} - \bar{x}).$$

Then, it may be shown that

$$E \delta_g^2 = \sigma^2 - \sigma_x^2,$$

and

$$E c_{k\mu g} = C_{k\mu} - \sigma_x^2.$$

Hence

$$E \delta_g^2 + s_g^2 \left(\frac{k-g}{g(k-1)} \right) = \sigma^2$$

¹¹ It may be wondered why we sample these S_i at random rather than systematically. If we sampled the S_i systematically, it would be equivalent to taking a single systematic sample having smaller intervals between elements of the sample. Furthermore, we could not derive the unbiased estimates of the sampling variance that we can now.

and

$$\mathfrak{E}\hat{c}_{k\mu g} + s_g^2 \left(\frac{k-g}{g(k-1)} \right) = C_{k\mu}.$$

The estimate, $r_{k\mu}$, of $\rho_{k\mu}$ defined by

$$\hat{c}_{k\mu g} + s_g^2 \left(\frac{k-g}{g(k-1)} \right) = r_{k\mu} \left[s_g^2 + s_g^2 \left(\frac{k-g}{g(k-1)} \right) \right]$$

is thus biased but in many cases the bias will be small. Of course, if $\mu = \frac{n}{2}$ when n is even then $r_{kn/2}$ is multiplied by 2 to estimate $\rho_{kn/2}$ as previously defined. Another approach would be to consider

$$gn_w s_g^2 = \sum_{\beta, \alpha} (x_{\beta+(\alpha-1)k} - \bar{x}_\beta)^2,$$

and

$$gn_w \hat{c}_{k\mu g} = \sum_{\beta, \alpha} (x_{\beta+(\alpha-1)k} - \bar{x}_\beta)(x_{\beta+(\alpha+\mu-1)k} - \bar{x}_\beta).$$

When that is done, it follows that

$$\mathfrak{E}_w s_g^2 = \sigma^2 - \sigma_z^2,$$

and

$$\mathfrak{E}_w \hat{c}_{k\mu g} = C_{k\mu} - \sigma_z^2,$$

and

$$\mathfrak{E}_w(s_g^2 + s_g^2) = \sigma^2,$$

$$\mathfrak{E}_w(\hat{c}_{k\mu g} + s_g^2) = C_{k\mu}.$$

Another estimate of $\rho_{k\mu}$ is thus defined by the equation

$$w\hat{c}_{k\mu g} + s_g^2 = w r_{k\mu} [w s_g^2 + s_g^2].$$

When $g = 1$, $s_g^2 = 0$ and we are unable to provide unbiased estimates of σ^2 , $C_{k\mu}$, and σ_z^2 from the sample. However, since

$$\frac{1 - r_{k\mu}}{1 - r_{k\mu'}} = \frac{s_g^2 - \hat{c}_{k\mu g}}{s_g^2 - \hat{c}_{k\mu' g}},$$

it follows that approximately

$$\frac{1 - \rho_{k\mu}}{1 - \rho_{k\mu'}} = \mathfrak{E} \frac{1 - r_{k\mu}}{1 - r_{k\mu'}},$$

since $\mathfrak{E}[s_g^2 - \hat{c}_{k\mu g}] = \sigma^2 - C_{k\mu}$. Similar equations hold for the $w r_{k\mu}$.

When we estimate the $\rho'_{k\delta}$, then the "within class" definition is simpler. Let $k(n - \delta)_{w\hat{c}'_{k\delta g}} = \sum_{i,\lambda} (x_{i+(\lambda-1)k} - {}_{1\delta}\bar{x}_i)(x_{i+(\lambda+\delta-1)k} - {}_{2\delta}\bar{x}_i)$, where

$$(n - \delta)_{1\delta}\bar{x}_i = \sum_{\lambda} x_{i+(\lambda-1)k},$$

$$(n - \delta)_{2\delta}\bar{x}_i = \sum_{\lambda} x_{i+(\lambda+\delta-1)k},$$

and let

$$g(n - \delta)_{w\hat{c}'_{k\delta g}} = \sum_{\beta,\lambda} (x_{\beta+(\lambda-1)k} - {}_{1\delta}\bar{x}_{\beta})(x_{\beta+(\lambda+\delta-1)k} - {}_{2\delta}\bar{x}_{\beta}).$$

Let

$$k\hat{c}'_{k\delta g} = \sum_i ({}_{1\delta}\bar{x}_i - \bar{x})({}_{2\delta}\bar{x}_i - \bar{x}),$$

and let

$$g\hat{c}_{k\delta g} = \sum_{\beta} ({}_{1\delta}\bar{x}_{\beta} - \bar{x})({}_{2\delta}\bar{x}_{\beta} - \bar{x}),$$

Then

$$\hat{c}'_{k\delta} = w\hat{c}'_{k\delta g} + b\hat{c}'_{k\delta g},$$

and

$$\mathfrak{E}(w\hat{c}'_{k\delta g} + b\hat{c}'_{k\delta g}) = C'_{k\delta}.$$

Thus, as estimates of the $\rho'_{k\delta}$ we obtain $r'_{k\delta}$ where

$$w\hat{c}'_{k\delta g} + b\hat{c}'_{k\delta g} = r'_{k\delta}(w\delta_g^2 + \delta_g^2).$$

In cases where \bar{x} is known simpler estimates of the $\rho_{k\mu}$, $\rho'_{k\delta}$, and σ^2 may be easily obtained since

$$\mathfrak{E} \sum_{\beta,\alpha} (x_{\beta+(\alpha-1)k} - \bar{x})(x_{\beta+(\alpha+\mu-1)k} - \bar{x}) = gnC'_{k\mu},$$

$$\mathfrak{E} \sum_{\beta,\lambda} (x_{\beta+(\lambda-1)k} - \bar{x})(x_{\beta+(\lambda+\delta-1)k} - \bar{x}) = g(n-\delta)f'_{1\delta},$$

and

$$\mathfrak{E} \sum_{\beta,\alpha} (x_{\beta+(\alpha-1)k} - \bar{x})^2 = gn\sigma^2.$$

Thus, in pilot studies, when \bar{x} is known it is possible to estimate the parameters in σ_x^2 from even a single sample.

9. Changes in the variance with changing size of sample. The chief reasons for expressing the variance of a systematic sampling design in terms of the variance of a random sample and the serial correlation coefficients were

1. To enable the making of comparisons with random and other sampling designs

2. To simplify the analysis of causes for the difference in the efficiencies of the systematic and random designs, and
3. To simplify the making of estimates of the variance for different sizes of sample.

In this section we are concerned with the third of these reasons. We shall discuss only the $\rho_{k\mu}$ since the analysis in terms of the $\rho'_{k\delta}$ is very similar.

The problem with which we are concerned is the estimation of the function, $\bar{\rho}_k$, of k . In order to show how this may be done for all values of k when the $\rho_{k\mu}$ have been computed for one value of k , let us first note that since σ^2 does not depend on k we may confine our considerations to the $C'_{k\mu}$. In section 6 we have defined $C_{k\mu}$ by the equation

$$knC_{k\mu} = \sum_r (x_r - \bar{x})(x_{r+k\mu} - \bar{x}).$$

Thus, if we wish to evaluate the $C_{k'\mu'}$ where k' is such that $k' \neq k$ and $k'n' = kn = N$, we have the result $C_{k'\mu'} = C_{k\mu}$ if $k'\mu' = k\mu$ and, thus, for any given values k' and μ' , we have

$$C_{k'\mu'} = C_{k'k'\mu'/k}$$

where we have replaced μ by $\frac{k'\mu'}{k}$.

This procedure will involve, if $k' < k$, some interpolation, but if the $\rho_{k\mu}$ are plotted against μ , this interpolation may often be carried through graphically. However, it is usually advisable to take k so that the possible values of k' are such that $k' > k$.

In some cases it may be possible to construct a correlation function. For example, if the x_r may be represented by a polynomial in r , then $\rho'_{k\delta}$ may be represented by a polynomial in δ . From that fact we conclude that if the x_r vary about a smooth trend the $\rho'_{k\delta}$ will also vary about a smooth trend and it may be possible to interpolate. Further investigation of this problem is necessary.

10. Stratified systematic sampling. In sampling practice it is customary to deal with stratified populations. The variance of an estimate based on a stratified population will usually not include the variability among the strata. Consequently, when a population is well stratified the variability of estimates based in a sample of size n will usually be considerably less than the variability of an estimate based on a random sample of size n , ignoring the strata. We now discuss the theory of systematic sampling from a stratified population.

Let us assume that the population, P , consists of L strata, P_1, \dots, P_L , the a th of which contains N_a elements x_{a1}, \dots, x_{aN_a} . It is desired to estimate the arithmetic mean, \bar{x} of P . Let the arithmetic mean of P_a be denoted by \bar{x}_a . Let $N_a = k_a n_a$.

We shall consider two possible cases, the first of which is often used because

of the administrative simplicity of giving identical operating instructions to the people selecting samples in different places. The results of this section will indicate when this method may be used.

Sampling Procedure I—Suppose that $k_1 = k_2 = \cdots k_L = k$, and that the sampling procedure consists in selecting one of the integers, $1, \cdots, k$ at random, each integer having a probability $1/k$ of being selected. Then, if the integer selected is, for example, i , the sample of P_a consists of $x_{a1}, x_{a1+k}, \cdots, x_{a1+(n_a-1)k}$. Thus, there are exactly k possible samples, S_1, \cdots, S_k , each having probability $1/k$ of being the actual sample obtained by performing the sampling procedure.

Sampling Procedure II—The sampling procedure consists in selecting one of the integers $1, \cdots, k_a$ at random, for each value of a , each integer having a probability of $1/k_a$ of being selected. Then, there are exactly $k_1 \cdots k_L$ possible samples, each having probability $1/k_1 \cdots k_L$ of being the actual sample obtained by performing the sampling procedure.

Other sampling procedures for stratified sampling, of course, exist. The two listed above, however, cover most practical problems except those involving cluster sampling. These will be treated in a later paper. Furthermore, from the conclusions derived concerning these procedures it will be possible to infer conclusions concerning other stratified sampling procedures.

Let S_{ai} be the class of elements $x_{a1}, x_{a1+k}, \cdots, x_{a1+(n_a-1)k}$. We consider sampling procedure I. A systematic sample of size n_a is to be selected from P_a . The possible samples are S_1, \cdots, S_k where S_i consists of all the elements in S_{1i}, \cdots, S_{Li} . Let the arithmetic mean of the elements in S_{ai} be denoted by \bar{x}_{ai} . Let the arithmetic mean of the sample from P_a be denoted by \bar{x}_a and let the sample mean be denoted by \bar{x} , where

$$N\bar{x} = N_1\bar{x}_1 + \cdots + N_L\bar{x}_L.$$

$$\text{Then } N\bar{x} = \sum_a N_a \bar{x}_a = \sum_a N_a \frac{1}{k} \sum_i \bar{x}_{ai} = N\bar{x}.$$

It follows from Appendix C, that

$$\sigma_{\bar{x}}^2 = \frac{1}{N^2} \sum_{a,b} N_a N_b \sigma_{\bar{x}_a \bar{x}_b}$$

where

$$\begin{aligned} \sigma_{\bar{x}_a \bar{x}_b} &= \bar{S}(\bar{x}_a - \bar{x}_a)(\bar{x}_b - \bar{x}_b) \\ &= \frac{1}{k} \sum_i (\bar{x}_{ai} - \bar{x}_a)(\bar{x}_{bi} - \bar{x}_b). \end{aligned}$$

Although the expression for $\sigma_{\bar{x}_a \bar{x}_b}$ can be further simplified, the important fact is that if corresponding items in different strata are positively correlated, it is inadvisable to use sampling procedure I unless other considerations than sampling error are dominant. But if the corresponding items are negatively correlated then sampling procedure I will yield a smaller variance than sampling procedure II.

We now consider sampling procedure II. The difference between sampling procedures I and II is that in sampling procedure II we know that $\sigma_{2_{a2b}} = 0$, if $a \neq b$ because of the separate selection of sample in each stratum. Thus, under sampling procedure II, $\sigma_2^2 = \frac{1}{N^2} \sum_a N^2 \sigma_{2a}^2$ where σ_{2a}^2 has been derived in section 6.

11. A comparison of the efficiencies of systematic and random sampling procedures. The study of any sampling technique is incomplete unless some comparisons are made with other possible sampling techniques. In this section the systematic sampling procedure is compared with the unrestricted random and stratified random sampling procedure.

The means and variances associated with the random and stratified random sampling procedures will be denoted by the use of primes (') and double primes (')') respectively.

Then we know that

$$\sigma_2^2 / \sigma_2'^2 = (1 + 2 \sum_{\mu} \rho_{k\mu}) \left(\frac{kn - 1}{kn - n} \right)$$

and consequently $\sigma_2^2 < \sigma_2'^2$ if

$$\sum_{\mu} \rho_{k\mu} < -(n - 1)/2(kn - 1).$$

If n is large relative to k , we may use $-\frac{1}{2}k$ as an approximation to $-(n - 1)/2(kn - 1)$.

In order to make more specific comparisons, it is useful to assume that the population elements x_r are given by some function of ν , and to assume some functions such as

$$x_r = A_0 + A_1 \nu + \cdots + A_h \nu^h,$$

or

$$\begin{aligned} x_r &= B_0 + A_1 \sin \frac{2\pi\nu}{N} + B_1 \cos \frac{2\pi\nu}{N} \\ &+ \cdots \\ &+ A_h \sin \frac{2\pi h\nu}{N} + B_h \cos \frac{2\pi h\nu}{N}, \end{aligned}$$

and then to investigate the efficiencies of the various possible sampling procedures on the bases of such assumed distributions of the x_r . It should be noted that the use of the systematic sampling technique involves the assumption that it is possible to order the elements of the population in a logical way, and then use this ordering in selecting the sample systematically.

We shall now consider several possibilities. Let us first note that if we are sampling but one element from a stratum then the variance of the stratum

sample mean is the same whether the sampling is random or systematic. On the other hand, it follows from section 10 that if we stratify the population into L strata so that a systematic sample of size L chooses the j th element of each stratum, say, then the variance of the mean of the stratified random sample will be greater or less than the variance of the mean of the systematic sample depending on whether the average correlation between strata sample means in the systematic sample is negative or positive.

Let us now consider the origin of the warnings against the use of systematic samples from a population having a periodic distribution. If k is the period, the correlation between the strata means of the systematic sample is $+1$ and hence the random sample is superior. However, if the period is $2k$ then we shall show that the systematic sample will probably have a smaller variance.

Suppose that the period is $2k$ and that within two adjoining strata of size k we always have $x_1 = x_{2k}, x_2 = x_{2k-1}, \dots, x_k = x_{k+1}$ and $x_1 - \bar{x} = -(x_{k+1} - \bar{x})$. Then, if we are sampling one element from each stratum, the correlation between the systematic sample means, (the individual elements in this case), will be -1 if the strata subscripts differ by an odd number and $+1$ if the strata subscripts differ by an even number.

The variance within each of the n strata is σ_1^2 , where

$$k\sigma_1^2 = \sum_{i=1}^k (x_i - \bar{x})^2.$$

The variance between strata means is zero. Hence $\sigma^2 = \sigma_1^2$. The variance of the mean of an unrestricted random sample of size n where $n = L$ is then $\sigma_{\bar{x}}^2 = \frac{N-L}{N-1} \frac{\sigma_1^2}{L}$ and the variance of the stratified random sampling mean is $\sigma_{\bar{x}}^2 = (1/L)\sigma_1^2$ while the variance of the systematic sampling mean is

$$\sigma_{\bar{x}}^2 = \frac{\sigma_1^2}{L^2} \sum_{i,j=1}^L (-1)^{i-j} (2 - \delta_{ij})$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.

Then it may be shown that if L is even $\sigma^2 = 0$ while, if L is odd $\sigma^2 = (1/L^2)\sigma_1^2$.

Consequently, the efficiency of the systematic sample mean is greater than the efficiency of the stratified random sample mean if the population has a periodic distribution and the size of stratum is half the period. It should be noted that the same situation holds for k equal to an even or odd multiple of half the period as held for k equal to the period of half the period.

The situation is quite different if we assume that the elements of the population have a straight line distribution. Without loss of generality, we may assume that the straight line distribution is given by $x_i = i$. Then for an unrestricted random sample of size n , the sample mean being denoted by \bar{x}' we have $E\bar{x}' = \bar{x} = \frac{1}{2}(kn + 1)$,

$$\sigma^2 = \frac{k^2 n^2 - 1}{12},$$

and

$$\sigma_{\bar{x}}^2 = \frac{(k-1)(kn+1)}{12}.$$

For a stratified random sampling design, let us assume that $N_1 = \dots = N_L = \frac{cN}{n}$ where c may equal any of the integers $1, 2, \dots, n$; i.e. $L = \frac{n}{c}$. Let $n_1 = \dots = n_L = c$. Then $\sigma_{\bar{x}''}^2 = \frac{c^2}{n^2} \frac{k-1}{ck-1} \sum_a \sigma_a^2$ where \bar{x}'' is the sample mean of the stratified random sample. If the a th stratum contains $x_{(a-1)cN/n+1}, \dots, x_{acN/n}$ then $\sigma_a^2 = \frac{c^2 k^2 - 1}{12}$ and $\sigma_{\bar{x}''}^2 = \frac{c}{n} \frac{k-1}{ck-1} \frac{c^3 k^2 - 1}{12}$. Finally

$$\begin{aligned} \sigma_{\bar{x}}^2 &= \sigma^2 - \frac{1}{kn^2} \sum_i \sum_j (x_i - x_{j+1+k})^2 \\ &= \sigma^2 - \frac{k^2(n^2-1)}{12} \\ &= \frac{k^2-1}{12}. \end{aligned}$$

To summarize

$$\begin{aligned} \sigma_{\bar{x}}^2 &= \frac{(k-1)(kn+1)}{12} = \frac{(k-1)(N+1)}{12}, \\ \sigma_{\bar{x}''}^2 &= \frac{c(k-1)(ck+1)}{12n} = \frac{(k-1)\left(\frac{N}{L}+1\right)}{12L}, \\ \sigma_{\bar{x}}^2 &= \frac{k^2-1}{12}. \end{aligned}$$

It is clear that both $\sigma_{\bar{x}}^2$ and $\sigma_{\bar{x}''}^2$ are less than $\sigma_{\bar{x}}^2$. However $\sigma_{\bar{x}}^2/\sigma_{\bar{x}''}^2 = \frac{L(k+1)}{\frac{N}{L}+1}$.

Since $kn = N$ and $cL = n$ it follows that $k = \frac{N}{cL}$ and hence $\sigma_{\bar{x}}^2 < \sigma_{\bar{x}''}^2$ if $N > L(L-1)$ and $c \geq \frac{L}{1 - \frac{L(L-1)}{N}}$. In all cases $\frac{N}{L} \geq c$. It follows therefore

that for a value of c to exist we must have $\frac{N}{L} > \frac{NL}{N - L(L-1)}$ as a result of which we find that N must exceed $2L^2 - L$. Hence $\sigma_{\bar{x}}^2 \leq \sigma_{\bar{x}''}^2$ if $N > 2L^2 - L$ and $c \geq \frac{L}{1 - \frac{L(L-1)}{N}}$. Otherwise $\sigma_{\bar{x}}^2 > \sigma_{\bar{x}''}^2$.

This result follows from two facts:

(1) If one element is being taken from each stratum then the high average correlation between strata means results in the efficiency of the stratified random sampling mean being greater than the efficiency of the systematic sampling mean, despite the equal within stratum variances.

(2) If more than one element is being taken from each stratum then the within stratum variance of the systematic sampling mean is less than the within stratum variance of the stratified random sampling mean and if the size of stratum and sample from stratum are large enough, the smaller within stratum variance of the systematic sample more than compensates for the correlation among strata means.

Of course, in a straight line distribution there are much more efficient methods of defining a stratified random sample than that we have used. Furthermore, more efficient sampling procedures than those discussed are available. However, this example will be of use in indicating the general problems that arise as well as the procedures that may be followed in attacking them.

Another comparison of systematic and stratified random sampling may be obtained by considering the x_r to be composed of two elements, a trend function and a periodic function so that the deviations from the trend constitute a periodic function

Let $x_r = \varphi_1(\nu) + \varphi_2(\nu)$, where $\varphi_1(\nu)$ is a trend function and $\varphi_2(\nu)$ is a periodic function of period $2h$, $N = 2hQ$.

Let $\varphi_2(\nu) = y_\nu$. Then $y_j = y_{2h+j} = \dots = y_{2h(Q-1)+j}$, $j = 1, \dots, 2h$ and $y_{2ha+j} - \bar{y} = -(y_{2h(a+1)+j} - \bar{y})$, $j = 1, \dots, h$, $a = 0, \dots, Q-1$.

Since the sizes of sample that we shall consider for purposes of this comparison are all multiples of h we shall calculate our variances and covariances so that we obtain all the necessary information at once.

Let the mean of $\varphi_1(\nu)$ be denoted by $\bar{\varphi}_1$ and let the mean of $\varphi_2(\nu)$ be denoted by \bar{y} . Then $\bar{x} = \bar{\varphi}_1 + \bar{y}$ and

$$\begin{aligned} N\sigma^2 &= \sum_r (x_r - \bar{x})^2 \\ &= \sum_r [\varphi_1(\nu) - \bar{\varphi}_1]^2 + \sum_r (y_r - \bar{y})^2 + 2 \sum_r [\varphi_1(\nu) - \bar{\varphi}_1](y_r - \bar{y}) \\ &= \sum_{a,1} (\varphi_{1(a-1)h+1} - \bar{\varphi}_{1a})^2 + h \sum_a (\bar{\varphi}_{1a} - \bar{\varphi}_1)^2 \\ &\quad + \sum_{a,1} (y_{(a-1)h+1} - \bar{y}_a)^2 + h \sum_a (\bar{y}_a - \bar{y})^2 \\ &\quad + \sum_{a,1} [\varphi_{1(a-1)h+1} - \bar{\varphi}_{1a}](y_{(a-1)h+1} - \bar{y}_a) + h \sum_a (\bar{\varphi}_{1a} - \bar{\varphi}_1)(\bar{y}_a - \bar{y}) \end{aligned}$$

where $a = 1, \dots, 2Q$; $i = 1, \dots, h$, $\bar{\varphi}_{1a}$ is the arithmetic mean of $\varphi_{1(a-1)h+1}, \dots, \varphi_{1ah}$ and \bar{y}_a is the arithmetic mean of $y_{(a-1)h+1}, \dots, y_{ah}$.

It follows from the assumptions with respect to the y_r that $\bar{y}_a = \bar{y}$ and that $\sum_i (y_i - \bar{y}_a)^2$ is the same for each value of a and is equal to $\sum_{i=1}^h (y_i - \bar{y})^2$.

Since $y_i = y_{2h+1} = \dots = y_{2h(Q-1)+1}$, and $y_{h+1} = y_{2h+1} = \dots = y_{2h(Q-1)+h+1}$ we have

$$\begin{aligned} \sum_{\nu} [\varphi_1(\nu) - \bar{\varphi}_1](y_{\nu} - \bar{y}) &= \sum_{i=1}^h (y_i - \bar{y}) \sum_{a=1}^Q [\varphi_1(2a-2)h+1 - \varphi_1(2a-2)] \\ &\quad + \sum_{i=h+1}^{2h} (y_i - \bar{y}) \sum_{a=1}^Q [\varphi_1(2a-2)h+1 - \varphi_1(2a-2)]. \end{aligned}$$

Since $y_i - \bar{y} = (y_{i+h} - \bar{y})$, we also have $\sum_{\nu} [\varphi_1(\nu) - \bar{\varphi}_1](y_{\nu} - \bar{y}) = \sum_{i=1}^h (y_i - \bar{y}) \left\{ \sum_{a=1}^Q [\varphi_1((2a-2)h+i) - \varphi_1((2a-1)h+i) - \varphi_1(2a-2) + \varphi_1(2a-1)] \right\}$ which vanishes for example if $\varphi_1(\nu)$ is a straight line or if $\varphi_1(\nu)$ is a succession of straight lines each having length $2h$.

Let us now assume that $\varphi_1(\nu) = A + B\nu$. Then $\bar{\varphi}_1 = A + B \left[(a-1)h + \frac{h+1}{2} \right]$,

$$\varphi_1((a-1)h+i) - \bar{\varphi}_1 = i - \frac{h+1}{2},$$

$$\sum_i [\varphi_1((a-1)h+i) - \bar{\varphi}_1]^2 = \frac{B^2 h(h^2-1)}{12},$$

$$\sum_a (\bar{\varphi}_1 - \bar{\varphi}_1)^2 = \frac{B^2 h^2 (4Q^2-1)(2Q)}{12},$$

$$\bar{\varphi}_1 = A + B \frac{2h+1}{2},$$

$$\sum_{\nu} (\varphi_1(\nu) - \bar{\varphi}_1)^2 = \frac{2hQB^2}{12} [4h^2Q^2 - 1].$$

Then $\sigma^2 = \sigma_y^2 + \frac{B^2}{12} [4h^2Q^2 - 1]$ where $h\sigma_y^2 = \sum_{\nu} (y_{\nu} - \bar{y})^2$, and the variance of the mean of an unrestricted random sample of size n is $\sigma_1^2 = \frac{N-n}{N-1} \sigma^2$.

Let us assume now that the size of stratum is mh where m is a factor of $2Q$, say $2Q/m = L_m$. Then the variance within each of the L_m strata is a constant, say, σ_1^2 where $\sigma_1^2 = \sigma_y^2 + \frac{B^2}{12} [h^2m^2 - 1]$ if m is even. If m is odd then L_m is even

and in half the strata the within stratum variance is $\sigma_1^2 + \frac{1}{hm} \sum_{i=1}^h (y_i - \bar{y}) \left(i - \frac{h+1}{2} \right)$ while in the other strata, the within stratum variance is $\sigma_1^2 - \frac{1}{hm} \sum_{i=1}^h (y_i - \bar{y}) \left(i - \frac{h+1}{2} \right)$.

Then, if c elements are sampled at random from each of the L_m strata, it follows that

$$\begin{aligned}\sigma_{2'}^2 &= \frac{1}{L_m} \left(\frac{mh - c}{mh - 1} \right) \frac{\sigma_1^2}{c} \\ &= \frac{1}{L_m} \left(\frac{mh - c}{mh - 1} \right) \frac{1}{c} \left(\sigma_y^2 + \frac{B^2}{12} [h^2 m^2 - 1] \right).\end{aligned}$$

In order to evaluate the variance of the systematic sampling mean let us evaluate $\sum_j (x_j - x_{j+ks})^2 = k(n - \delta)s'_{ks}$. Now upon substituting for x_j , it follows that $k(n - \delta)s'_{ks} = \sum_j (y_j - y_{j+ks})^2 - 2Bk\delta \sum_j (y_j - y_{j+ks}) + k(n - \delta)B^2k^2\delta^2$.

Then, if k is a multiple of h , it follows that $\sum_j (y_j - y_{j+ks}) = 0$. Furthermore, if k is an even multiple of h , then $y_j = y_{j+ks}$ and hence $\sum_j (y_j - y_{j+ks})^2 = 0$. Finally, if k is an odd multiple of h then, if δ is an odd number $y_j - y_{j+ks} = 2(y_j - \bar{y})$ while, if δ is even $y_j - y_{j+ks} = 0$ and hence

$$\begin{aligned}\sum_j (y_j - y_{j+ks})^2 &= 4 \sum_j (y_j - \bar{y})^2 \\ &= 4 \frac{k(n - \delta)}{h} \sum_j (y_j - \bar{y})^2\end{aligned}$$

if k is an odd multiple of h and δ is an odd number. Note that if k is an odd multiple of h , then n is an even number. Since

$$\sigma_2^2 = \sigma^2 - \frac{1}{n^2} \sum_j (n - \delta)s'_{ks}$$

it is necessary to evaluate $\sum_j (n - \delta)s'_{ks}$. Now, if k is an even multiple of h , it follows that $(n - \delta)s'_{ks} = (n - \delta)B^2k^2\delta^2$ and

$$\begin{aligned}\sum_j (n - \delta)s'_{ks} &= B^2k^2 \{ n \sum_j \delta^2 - \sum_j \delta^3 \} \\ &= B^2k^2 \frac{n^3(n^2 - 1)}{12}\end{aligned}$$

Hence, if k is an even multiple of h , it follows that $\sigma_2^2 = \sigma^2 - \frac{B^2k^2(n^2 - 1)}{12}$.

On the other hand, if k is an odd multiple of h , and if δ is odd, we have $(n - \delta)s'_{ks} = (n - \delta)B^2k^2\delta^2 + 4(n - \delta)\sigma_y^2$ while if δ is even $(n - \delta)s'_{ks} = (n - \delta)B^2k^2\delta^2$.

Hence

$$\sum_j (n - \delta)s'_{ks} = \frac{B^2k^2n^2(n^2 - 1)}{12} + n^2\sigma_y^2.$$

Hence, if k is an odd multiple of h , it follows that

$$\sigma_z^2 = \sigma^2 - \frac{B^2 k^2 (n^2 - 1)}{12} = \sigma^2.$$

Then, if k is an even multiple of h

$$\begin{aligned}\sigma_z^2 &= \sigma_v^2 + \frac{B^2}{12} (k^2 n^2 - 1) - \frac{B^2}{12} k^2 (n^2 - 1) \\ &= \sigma_v^2 + \frac{B^2}{12} (k^2 - 1),\end{aligned}$$

and, if k is an odd multiple of h , then $\sigma_z^2 = \frac{B^2}{12} B^2 (k^2 - 1)$.

Thus, systematic sampling will yield superior results if

$$c > \frac{\frac{L}{12\sigma_v^2}}{1 + \frac{B^2(hm)(hm-1)}{12\sigma_v^2}} = \frac{L(L-1)}{N}.$$

Since $\frac{N}{L} > c$ it follows that for a solution, c , to exist, we must have

$$N > 2L^2 - L - \frac{12\sigma_v^2}{B^2} \left(\frac{L^2}{N} - L \right).$$

12. Summary. In this paper we have presented the theoretical basis for systematic sampling for stratified and unstratified populations including the derivation of the variances, a study of the possible values of the parameters, estimates of the parameters, the effects of changing the size of sample, and comparisons among systematic sampling, unrestricted random sampling, and stratified random sampling. The paper contains for the case where the sampling unit consists of one element, not only the theory necessary, but in addition, some analysis of the conditions under which systematic sampling ought be used, and formulas for calculating the variances.

In later papers of this series, we shall present the theory of systematic sampling when the sampling unit is a cluster of elements, the theory when we assume we are sampling not from a finite population but an infinite population, each of whose elements is normally distributed, and further studies of various parts of the theory and practice of systematic sampling.

APPENDIX A

Values Assumed by Certain Variables

In order to avoid repeating the limits of summation of variables, we shall give these limits in this appendix.

TABLE I
Values Assumed by Subscripts

Letter	The letter will assume all integral values from 1 to
i	k
λ	$n - \delta$
j	$k(n - \delta)$
δ	$n - 1$
ν, ν'	kn
α	n
γ	n
μ, μ'	$n/2$ if n is even, $\frac{n-1}{2}$ if n is odd
a, b	L

The letter β will assume the values i_1, i_2, \dots, i_g where i_1, \dots, i_g are a selection of g of the k integers $1, \dots, k$.

APPENDIX B

On the Limits of Some Finite Sums

The difficulties that arise in the transformation of finite sums are very similar to those that arise in the theory of transforming multiple integrals, i.e., the effects of transforming variables or order of summation on the limits of summation. Certain lemmas that have proved useful in this paper are presented separately here in a more general form.

Let $f(u)$ and $f(u, v)$ be functions of u and v that are finite for all possible values of u and v .

LEMMA 1.

$$\sum_{\substack{\alpha, \gamma \\ \alpha < \gamma}} f(x_{i+(\alpha-1)k}, x_{i+(\gamma-1)k}) = \sum_{\delta, \lambda} f(x_{i+(\lambda-1)k}, x_{i+(\lambda+\delta-1)k})$$

PROOF: Let $\alpha = \lambda$ and let $\gamma = \lambda + \delta$. Since $1 \leq \alpha < \gamma$ and $\gamma \leq n$, the possible values of δ are $1, \dots, (n-1)$. For any fixed value of δ the possible values of λ then are 1 to $n - \delta$ since $\lambda = \gamma - \delta$ and, for a fixed value of δ the maximum value of λ is determined when $\gamma = n$. With these limits each term of f on the left side of the equation occurs once and only once on the right side of the equation. Furthermore, no additional term occurs on the right side of the equation.

LEMMA 2.

$$\sum_i \sum_{\lambda} f(x_{i+(\lambda-1)k}, x_{i+(\lambda+\delta-1)k}) = \sum_i f(x_i, x_{i+\delta k}).$$

PROOF: Let $j = i + (\lambda - 1)k$. Then j is a monotone increasing function of i and λ . The minimum value of j occurs when $i = 1$. In that case $j = 1$. The maximum value of j occurs when $i = k, \lambda = n - \delta$. In that case $j = (n - \delta)k$.

With these limits each term of f on the left side of the equation occurs once and only once on the right side of the equation. Furthermore, no additional term occurs on the right side of the equation.

LEMMA 3.

$$\sum_{\substack{i, \alpha, \gamma \\ \alpha < \gamma}} f(x_{i+(\alpha-1)k}, x_{i+(\gamma-1)k}) = \sum_{\delta, i} f(x_i, x_{i+k\delta}).$$

PROOF: First apply Lemma 1 to $\sum_{\substack{\alpha, \gamma \\ \alpha < \gamma}} f(x_{i+(\alpha-1)k}, x_{i+(\gamma-1)k})$ and then apply Lemma

2 to the resulting expression.

LEMMA 4.

$$\sum_{\delta, i} [f(x_i) + f(x_{i+k\delta})] = (n-1) \sum_r f(x_r).$$

PROOF: Let $m = j + k\delta$. Then for any fixed value of δ the minimum value of m occurs when $j = 1$. In that case $m = k\delta + 1$. For any fixed value of δ , the maximum value of m occurs when $j = k(n - \delta)$. In that case $m = nk$. The letter m will assume all integral values from $k\delta + 1$ to kn , and hence,

$$\sum_{\delta, i} f(x_i) + \sum_{\delta, i} f(x_{i+k\delta}) = \sum_{\delta, i} f(x_i) + \sum_{\delta, m} f(x_m).$$

If we sum δ from $n-1$ to 1 instead of from 1 to $n-1$ in $\sum_{\delta, m} f(x_m)$ we see that

$$\begin{aligned} \sum_{\delta, i} f(x_i) + \sum_{\delta, m} f(x_m) &= \sum_{j=1}^{k(n-1)} f(x_j) + \sum_{m=k(n-1)+1}^{kn} f(x_m) \\ &+ \dots \\ &+ \sum_{j=1}^k f(x_j) + \sum_{m=k+1}^{kn} f(x_m) \end{aligned}$$

where the summations of x_j are terms of $\sum_{\delta, i} f(x_i)$ and the summations of x_m are terms of $\sum_{\delta, m} f(x_m)$. But $\sum_{j=1}^{k(n-\delta)} f(x_j) + \sum_{m=k(n-\delta)+1}^{kn} f(x_m) = \sum_r f(x_r)$ and hence Lemma 4 is proved.

LEMMA 5. Let

$$\sum_{i, \alpha, \gamma} f(x_{i+(\alpha-1)k}) f(x_{i+(\gamma-1)k}) = A.$$

Then

$$A = n \sum_v [f(x_v)]^2 - \sum_{i, \delta} [f(x_i) - f(x_{i+k\delta})]^2.$$

PROOF:

$$A = \sum_{i, \alpha} [f(x_{i+(\alpha-1)k})]^2 + 2 \sum_{\substack{i, \alpha, \gamma \\ \alpha < \gamma}} f(x_{i+(\alpha-1)k}) f(x_{i+(\gamma-1)k}).$$

By Lemma 3

$$2 \sum_{\substack{i, \alpha, \gamma \\ \alpha < \gamma}} f(x_{i+(a-1)k}) f(x_{i+(\gamma-1)k}) = 2 \sum_{i, j} f(x_i) f(x_{j+k}),$$

and since we have

$$2f(x_j)f(x_{j+k}) = f(x_j)^2 + f(x_{j+k})^2 - [f(x_j) - f(x_{j+k})]^2,$$

the proof is completed by using Lemma 4.

LEMMA 6. Let $kn\bar{f} = \sum_v f(x_v)$. Then

$$A\left(\frac{1}{kn^2}\right) - \bar{f}^2 = \left(\frac{1}{kn}\right) \sum_v [f(x_v) - \bar{f}] - \left(\frac{1}{kn^2}\right) \sum_{i, b} [f(x_i) - f(x_{i+kb})]^2.$$

PROOF: This lemma is a direct consequence of Lemma 5.

LEMMA 7.

$$A = \sum_v [f(x_v) - \bar{f}]^2 + 2 \sum_{i, b} [f(x_i) - \bar{f}][f(x_{i+kb}) - \bar{f}] + kn^2 \bar{f}^2.$$

PROOF: From Lemma 4, it follows that

$$A = n \sum_v f(x_v)^2 - \sum_{i, b} \{[f(x_i) - \bar{f}]^2 + [f(x_{i+kb}) - \bar{f}]^2 + 2 \sum_{i, b} [f(x_i) - \bar{f}][f(x_{i+kb}) - \bar{f}]\}$$

and hence, from Lemma 3, it follows that

$$A = n \sum_v [f(x_v) - \bar{f}]^2 + n^2 k \bar{f}^2 - (n-1) \sum_v [f(x_v) - \bar{f}]^2 + 2 \sum_{i, b} [f(x_i) - \bar{f}][f(x_{i+kb}) - \bar{f}],$$

whence the lemma is proved.

LEMMA 8.

$$A\left(\frac{1}{kn^2}\right) - \bar{f}^2 = \left(\frac{1}{kn^2}\right) \sum_v [f(x_v) - \bar{f}]^2 + \left(\frac{2}{kn^2}\right) \sum_{i, b} [f(x_i) - \bar{f}][f(x_{i+kb}) - \bar{f}].$$

This lemma is a direct consequence of Lemma 7.

LEMMA 9. If $h > kn$, let x_h equal x_{h-kn} . Let $f(u, v) = f(v, u)$ i.e. f is symmetric. Then, if we let

$$d_{kb} = \sum_i f(x_i, x_{i+kb})$$

it follows that

$$d_{kb} + d_{kn-b} = \sum_v f(x_v, x_{v+kb}).$$

PROOF: Obviously

$$\sum_v f(x_v, x_{v+kb}) = d_{kb} + B,$$

where

$$B = \sum_{g=k(n-\delta)+1}^{kn} f(x_g, x_{g+kd}).$$

Now, let $h = g - (n - \delta)k$. Then

$$B = \sum_{h=1}^{hk} f(x_{h+(n-\delta)k}, x_{h+kn}).$$

Since $x_{h+kn} = x_h$, and $f(x_{h+(n-\delta)k}, x_h) = f(x_h, x_{h+(n-\delta)k})$, it follows that $B = d_{kn-\delta}$ and the lemma is proved. It is noted that the symmetry of $f(u, v)$ is necessary as well as sufficient, for if $f(x_r, x_{r+kd}) = x_r - x_{r+kd}$ the theorem is false.

APPENDIX C

Stratified Sampling

Let the population P consist of L strata P_1, \dots, P_L . Let \bar{x} be the arithmetic mean of P , and \bar{x}_a the arithmetic mean of P_a . Let \tilde{x}_a be the sample estimate of \bar{x}_a , and let $\tilde{x} = \sum_a c_a \tilde{x}_a$. Then $\mathcal{S}\tilde{x} = \sum_a c_a A_a = A$ where $\mathcal{S}\tilde{x}_a = A_a$. Let $\sigma_{\tilde{x}}^2$ be defined by $\sigma_{\tilde{x}}^2 = \mathcal{S}(\tilde{x} - \bar{x})^2$. Then $\sigma_{\tilde{x}}^2 = \mathcal{S}(\tilde{x} - A)^2 + (A - \bar{x})^2$ and hence it is easy to see that $\sigma_{\tilde{x}}^2 = \sum_{a,b} c_a c_b \sigma_{\tilde{x}_a \tilde{x}_b} + (A - \bar{x})^2$ where

$$\sigma_{\tilde{x}_a \tilde{x}_b} = \mathcal{S}(\tilde{x}_a - A_a)(\tilde{x}_b - A_b),$$

$$(A - \bar{x})^2 = \left[\sum_a \left(c_a A_a - \frac{N_a}{N} \bar{x}_a \right) \right]^2,$$

and if $NC_a = N_a$, then

$$(A - \bar{x})^2 = \sum_{a,b} c_a c_b (A_a - \bar{x}_a)(A_b - \bar{x}_b)$$

$$\text{and } \sigma_{\tilde{x}}^2 = \sum_{a,b} c_a c_b \sigma_{\tilde{x}_a \tilde{x}_b}$$

where

$$\sigma_{\tilde{x}_a \tilde{x}_b} = \mathcal{S}(\tilde{x}_a - \bar{x}_a)(\tilde{x}_b - \bar{x}_b).$$

These formulae hold whatever may be the method used in sampling the i th stratum. If \tilde{x} is an unbiased estimate of \bar{x} and \tilde{x}_a is independent of \tilde{x}_b , then the usual formula $\sigma_{\tilde{x}}^2 = \sum_a c_a^2 \sigma_{\tilde{x}_a}^2$ holds. The formula for $\sigma_{\tilde{x}_a \tilde{x}_b}$ will, of course, depend on whether a random, cluster, systematic, or other sampling procedure is used.

ON THE PROBABILITY THEORY OF LINKAGE IN MENDELIAN HEREDITY

BY HILDA GEIRINGER

Bryn Mawr College

1. **Introduction.** If for a certain generation the distribution of genotypes is known and a certain law of heredity is assumed, the distribution of genotypes in the next generation can be computed. Suppose there are N different genotypes in the n th generation in the proportions $x_1^{(n)}, \dots, x_N^{(n)}$ where $\sum_{i=1}^N x_i^{(n)} = 1$ and denote by $p_{\kappa\lambda}^i$ the probability that an offspring of two parents of types κ and λ be of type i where $\sum_{i=1}^N p_{\kappa\lambda}^i = 1$ for all κ and λ , and $p_{\kappa\lambda}^i = p_{\lambda\kappa}^i$. Assuming panmixia, identical distributions $x_i^{(n)}$ for males and females, etc., we can derive $x_i^{(n+1)}$ from $x_i^{(n)}$ by means of the formula

$$(1) \quad x_i^{(n+1)} = \sum_{\kappa, \lambda=1}^N p_{\kappa\lambda}^i x_{\kappa}^{(n)} x_{\lambda}^{(n)} \quad (i = 1, 2, \dots, N).$$

Thus if the distribution $x_i^{(0)}$ is given for an initial generation we can deduce successively the $x_i^{(1)}, x_i^{(2)}, \dots$ for subsequent generations. Besides, one may wish to express the $x_i^{(n)}$, for any n , explicitly in terms of the initial distribution $x_i^{(0)}$, i.e. to "solve" the system (1). A further problem consists in determining the limit-distribution of the genotypes $\lim_{n \rightarrow \infty} x_i^{(n)}$ ($i = 1, \dots, N$).

Mendel's heredity theory is based on some ingenious assumptions which are known as Mendel's first and second law. They enable us to define the possible genotypes and to establish the recurrence formula (1); they will be explained and formulated in sections 2 and 3. It is well known that in Mendel's theory it makes an essential difference whether one or more "Mendelian characters" are considered. In the first case Mendel's first law only is used; there are with respect to this character but $N = 3$ different types and the recurrence formula (1) can be derived without difficulty. As early as 1908 (J. H. Hardy [5]) established the simple but most remarkable result that under random breeding a state of equilibrium is reached in the first filial generation, i.e. $x_i^{(1)} \neq x_i^{(0)}$ (in general) but $x_i^{(n)} = x_i^{(1)}$ ($n = 2, 3, \dots$).¹

In the case of $m \geq 2$ Mendelian characters Mendel assumed independent assortment of these characters (Mendel's second law). However, within four years after the dramatic rediscovery of Mendel's fundamental paper [10], observations were reported that did not show the results expected for two independent characters. T. H. Morgan [11] and collaborators in basic contributions, con-

¹ See also [12] where the stability of the particular ratio 1:2:1 is recognized.

cluded that a certain *linkage* of genes was to be assumed.² Taking that as the starting point, the main purpose of this paper is to establish the basic recurrence formula for the general case of linkage, to solve the corresponding system of difference equations, and to determine the limit distribution of genotypes. Throughout the paper "multiple alleles" are considered instead of making the frequent restriction to two alleles. This generalization is, however, an obvious one (see section 1).

In order to deal with the general problem a *linkage distribution* (l.d.), is introduced. This concept, which seems to be basic to the whole problem, refers to the probability theory of arbitrarily linked events [3]. The *crossover probabilities*, (c.p.), defined by Morgan and Haldane are, notwithstanding their high importance, not sufficient for our purpose. (They turn out to be certain marginal distributions of the l.d.) If, however, $m = 2$ and $N = 10$ (for two alleles), a case studied by W. Weinberg [16] H. S. Jennings [7] and R. B. Robbins [14], the c.p. is equivalent to the l.d. But for the general case the l.d. is needed and the desired results must be derived by other methods than explicit computation, which is feasible if m equals one or two. The original problem of independent assortment appears, of course, as a particular case of the general linkage. This problem was completely solved in 1923 by H. Tietze [15] in a very interesting but rather involved paper. The proof of the limit theorem given in the following pages for the general case is far simpler and shorter than the treatment of the particular case in the older paper and is therefore a simpler proof of Tietze's theorem.

After a brief consideration of the classical case $m = 1$ (section 2), the problem of m arbitrarily linked characters is discussed in section 3 with a particular view to a clear statement of the biological and mathematical assumptions. The l.d., its relation to the c.p., and some basic properties of both are considered in section 4. Then, after a very concise consideration of the case $m = 2$ (section 5), the basic recurrence formula is established in section 6 from which we deduce in section 7 two general limit theorems. The main point is that the limit distribution of genotypes is "uncorrelated" and equals the product of certain marginal distributions of first order deduced from the distribution for the first filial generation. As a kind of an appendix section 8 contains the solution of the system of equations furnished by the general recurrence formula.

In the second part of the paper an attempt is made to contribute to the *linear theory* or theory of the linear order of the genes, from the point of view of probability theory. Accordingly, the linear theory consists in certain assumptions on the l.d., or on an equivalent distribution which will be called *crossover distribution*, (c.d.), and which is more appropriate for this purpose. (Sections 9

² "To T. H. Morgan and his associates and students is due the credit for opening up this new field of genetic research; and the small vinegar fly *Drosophyla Melanogaster* upon which most of their work has been based, has now assumed as great an importance in genetics as the famous peas studied by Mendel." (Sinnot and Dunn, *Principles of Genetics*, New York 1939.)

and 10.) In this connection in section 10 a probability definition of the "distance" d_{ij} of two genes is proposed which, far from being contradictory to Morgan's ideas on the subject seems to formulate them mathematically; the distance d_{ij} between two genes i and j is defined as the mathematical expectation of the number of crossovers between i and j). This distance is of course additive as it ought to be in the framework of the linear theory.

A problem frequently discussed is whether the crossover probabilities are independent of each other (this independence is not identical with Mendel's free assortment). Observations (see [4a]) did not seem to substantiate this as a general assumption. Then it was concluded that there exists a so-called *interference* which prevents, i.e. diminishes the probability of crossovers too "near" to each other. (See also [13].) It seems to the author that observations on interference should be interpreted in terms of appropriate assumptions regarding the l.d. or the c.d. Again the remark holds that the c.p.'s are not sufficient for describing the situation. Hence in section 11 an attempt is made to understand "interference" by means of the c.d., accepting however the linear theory. It is well known that the explicit presentation of consistent dependent distributions is not trivial (see e.g. [2]). Not many different types of "contagious" distributions are known. In section 11 two such schemes are proposed which, though simple enough, seem to correspond to the general idea of interference. They contain as particular cases the case of independent and the case of disjoint crossovers.

2. One Mendelian character. Hardy's theorem. It will be helpful to start with the simple and well known case of one character introducing the basic concepts in a way appropriate for generalization.

Mendel recognized that the distribution of certain hereditary attributes in organisms is similar to the distribution of attributes in a probability distribution. With respect to a *Mendelian character* each individual is characterized by two elements called *genes* which represent two possible alternatives. The color of the flower of peas is such a character, the alternatives being red and white. With respect to this character each plant belongs to one of the three types: red-red, red-white, white-white.³ These are three different *genotypes*. In this paper genotypes only will be considered. The difference between genotypes and phenotypes and the related concepts of dominant and recessive qualities will not be dealt with. This is an example of a *two-valued* Mendelian character, i.e. a character for which only two possibilities exist or, using a more technical term,

³ It will be assumed throughout that the individuals considered are "diploid". That means in the terminology of the preceding example that the only possible types are RR, RW, and WW; or, using A and a: AA, Aa, and aa. Modern research has however revealed that situations may arise where "tetraploids", "hexaploids", etc. briefly "polyploids" prevail, i.e. types like $A^x a^y$ (with $x + y = 2p$). In this case the reproduction cell aggregates $A^x a^y$ (with $x_1 + y_1 = p$). Stability is no longer reached in the first filial generation. See [4b].

with two alleles. The case of two alleles is most frequently considered in the biological literature where the two possibilities correspond mostly to a dominant and a recessive quality. There is, however, no difficulty in considering from the very beginning the general case of *multiple alleles* where the character under consideration is assumed to be r -valued, i.e. susceptible to r different manifestations (e.g. $r = 5$ possible colors of a plant). These r possible values may be distinguished by the r arguments, $1, 2, \dots, r$.⁴

In the consideration of only one Mendelian character *Mendel's first law* only is used which may be stated as follows:

(a). With respect to one r -valued Mendelian character each individual belongs to one of the $r(r+1)/2$ possible types, each type being determined by a pair of elements (genes) x and y $\left(\begin{matrix} x = 1, \dots, r \\ y = 1, \dots, r \end{matrix} \right)$.

(b). In the formation of a new individual each parent transmits one of its two genes to the new individual, the other gene coming from the other parent.

(c). The probability for the transmission of either gene is the same and thus equals $\frac{1}{2}$.

We wish to deduce the distribution of genotypes in the $(n+1)$ st generation from the distribution of genotypes in the n th generation *under the assumption of complete panmixia* (random breeding). Moreover, assume that the given initial distributions of genotypes as well as the laws of heredity are the same for males and females.⁵ In computing successively the new distribution from the preceding one we shall always assume that the distribution of individuals participating in the process of procreation is the same as their distribution when born.

Let us denote a genotype by $(x; y)$, ($x = 1, \dots, r; y = 1, \dots, r$). To fix the ideas we shall assume through this paper that the gene x before the semicolon was transmitted by the mother, and the y after the semicolon by the father of the individual. In some cases which will be considered later this distinction will be relevant. Denote by $w^{(n)}(x; y)$ the probability of the type $(x; y)$ in the n th generation. Since the laws of heredity are the same for males and females we have $w^{(n)}(x; y) = w^{(n)}(y; x)$ and thus have for each generation a symmetric distribution of genotypes with r^2 probabilities whose sum is one. There is, however, according to principle (a) no difference between the types $(x; y)$ and $(y; x)$ and therefore it is preferable to group together these types, thus introducing for $x = 1, \dots, r; y = 1, \dots, r$:

$$\begin{aligned} (2) \quad & v^{(n)}(x; x) = w^{(n)}(x; x) \\ & v^{(n)}(x; y) = w^{(n)}(x; y) + w^{(n)}(y; x) \text{ where } x < y. \end{aligned}$$

⁴"It is simplest to deal with mere pairs of alternative conditions (alleles) but a theory remains seriously inadequate unless capable of extension to multiple alleles." ([17] p. 224).

⁵It is obvious that we may admit without any change of result different distributions for males and females in the initial generation, as long as random mating takes place afterwards.

Consequently there are $r(r+1)/2$ such probabilities:

$$(2') \quad v^{(n)}(1; 1), v^{(n)}(1; 2), \dots, v^{(n)}(1; r), v^{(n)}(2; 2), \dots, v^{(n)}(2; r), \dots, v^{(n)}(r; r),$$

where

$$(3) \quad \sum_{x=y} v^{(n)}(x; y) = 1 \quad (n = 0, 1, 2, \dots)$$

Now define $p^{(n)}(x)$ as the probability that in the n th generation a male (or a female) individual transmits the gene x . Obviously we have:

$$(4) \quad p^{(n)}(x) = \frac{1}{2} v^{(n)}(1; x) + \frac{1}{2} v^{(n)}(2; x) + \dots + v^{(n)}(x; x) \\ + \frac{1}{2} v^{(n)}(x; x+1) + \dots + \frac{1}{2} v^{(n)}(x; r)$$

and

$$(4') \quad \sum_{x=1}^r p^{(n)}(x) = 1$$

In fact, the gene x will be transmitted, if an individual possesses this gene and also transmits it. The individuals of type $(y; x)$ (or $(x; y)$) all possess the gene x and transmit it with probability $\frac{1}{2}$ if $y \neq x$ and with probability 1 if $y = x$. Besides, the probability of the type $(x; y)$ in the $(n+1)$ st generation is obviously $p^{(n)}(x)p^{(n)}(y)$:

$$(5) \quad w^{(n+1)}(x; y) = p^{(n)}(x)p^{(n)}(y) = w^{(n+1)}(y; x)$$

or in terms of the $v^{(n)}(x; y)$

$$(5') \quad v^{(n+1)}(x; x) = [p^{(n)}(x)]^2 \\ v^{(n+1)}(x; y) = 2p^{(n)}(x)p^{(n)}(y) \quad (x \neq y).$$

Hence by (4) and (5'), $v^{(n+1)}$ has been expressed in terms of $v^{(n)}$ and the recurrence-problem is solved. The distribution $w^{(n+1)}(x; y) (n \geq 0)$ shows "independence," and is therefore known to be stable. In fact, computing in the same way $p^{(n+1)}(x)$ we get

$$p^{(n+1)}(x) = \frac{1}{2} \cdot 2p^{(n)}(1)p^{(n)}(x) + \dots + p^{(n)}(x)p^{(n)}(x) \\ + \frac{1}{2} \cdot 2p^{(n)}(x)p^{(n)}(x+1) + \dots + \frac{1}{2} \cdot 2p^{(n)}(x)p^{(n)}(r) \\ = p^{(n)}(x) \cdot \sum_{\rho=1}^r p^{(n)}(\rho) = p^{(n)}(x)$$

or

$$(6) \quad p^{(n+1)}(x) = p^{(n)}(x) \\ (n = 0, 1, 2, \dots), (x = 1, 2, \dots, r).$$

This last formula contains G. H. Hardy's famous result [5] that $p^{(n)}(x)$ is the same for all n :

$$(7) \quad p^{(n)}(x) = p^{(0)}(x) \quad (n = 1, 2, \dots,)$$

and because of (5'):

$$(7') \quad v^{(n)}(x; y) = v^{(1)}(x; y) \quad (x \neq y, n = 2, 3, \dots).$$

In case of only one Mendelian property the distribution of genotypes reaches a stationary state in the first filial generation.

3. Basic assumptions in case of m Mendelian characters. A new situation presents itself if there is more than one character. In case of m characters a genotype is described by $2m$ numbers ($x_1, \dots, x_m; y_1, \dots, y_m$) or briefly $(x; y)$ (e.g. for $m = 5, r = 9: (1, 2, 3, 4, 6; 2, 7, 3, 5, 9)$). There are primarily $N = r^{2m}$ possible types because on each of the $2m$ places any of the r numbers can be written. Now, if the types (x, y) and (y, x) are considered as identical genotypes, the number of different genotypes reduces to $N_1 = \frac{r^m}{2}(r^m + 1)$ (e.g. for $r = 2, m = 1: N_1 = 3$; for $r = m = 2: N_1 = 10$). It is essential for the understanding of linkage that in counting this way two types like $(1, 3; 5, 7)$ and $(1, 7; 5, 3)$ or $(1, 1; 2, 2)$ and $(1, 2, 2, 1)$ are considered as different although in both cases the individual possesses with respect to the first character the gene pair 1, 5 and with respect to the second the pair 3, 7. If no difference is assumed between two such types the number of different genotypes reduces to $N_2 = \left(\frac{r(r+1)}{2}\right)^m$. (E.g. for $r = 2: N = 4^m, N_1 = \frac{1}{2} \cdot 2^m(2^m + 1), N_2 = 3^m$; hence for $m = r = 2: N = 16, N_1 = 10, N_2 = 9$ or for $r = 2, m = 3: N = 64, N_1 = 36, N_2 = 27$). Which method of counting is the correct one?

The answer is that there are but N_2 different genotypes if *Mendel's second law*, the *law of independent assortment*, is accepted. Then and only then there is no difference between types like $(1, 3; 5, 7)$ and $(1, 7; 5, 3)$. Under the assumption of general linkage however, these types must be distinguished, not as individuals, but with respect to their heredity properties, i.e. considered as parents of a new generation. Under this assumption there are in general N_1 different types. This will be discussed presently in more detail.

Let us first consider Mendel's original theory as contained in his *first and second law*. Analogous to (a), (b) and (c) in §2 we now formulate as follows:

(a') With respect to m characters the genotype of an individual is characterized by m pairs of numbers. Two individuals are of the same type if to each of the m characters corresponds the same pair. Hence there are $N_2 = \left(\frac{r(r+1)}{2}\right)^m$ genotypes.

(b') In the formation of a new individual a parent of type $(x_1, \dots, x_m; y_1, \dots, y_m)$ transmits to the offspring, corresponding to each of the m characters, one of the two genes which he (or she) possesses with respect to this character.

(c') The probability of transmitting any of these 2^m combinations is the same and therefore equal to $1/2^m$.

Consider e.g. the individual $(1,2,3;1,4,7)$; the pair 1,1 corresponds to the first character the pair 2,4 to the second and 3,7 to the third. Under the assumptions of Mendel's original theory this individual is of the same type with $(1,4,3;1,2,7)$ and $(1,2,7;1,4,3)$, and of course with $(1,4,7;1,2,3)$, etc. As $m = 3$, it may transmit eight combinations which in the preceding example reduce to four, because the individual is homozygous in the first character. These four combinations are 1,2,3 or 1,4,3 or 1,2,7 or 1,4,7 each with probability $2 \times \frac{1}{8} = \frac{1}{4}$.

The distribution of genotypes in successive generations under the assumption of Mendel's second law has been investigated by H. Tietze [15] who also considers the limiting distribution as $n \rightarrow \infty$. His results will appear as a particular case of our general considerations.

In order to discuss the basic facts which lead to the idea of linkage let us for the moment consider the case $m = 2$. Soon after the rediscovery of Mendel's work Bateson and Punnett reported observations which did not give the expected numerical results. To understand the type of such an observation assume that a homozygous male of type $(1,1;1,1)$, [or any other homozygous type, e.g. $(2,3;2,3)$] is mated to a homozygous female of type $(2,2;2,2)$ [or to any homozygous type different from the first e.g. $(4,5;4,5)$]. Obviously, in this case there is only one possible kind of offspring namely $(2,2;1,1)$, [or $(4,5;2,3)$]. But if now one of these daughters is mated to a homozygous male of the original type $(1,1;1,1)$, there are four kinds of possible offspring, namely $(2,2;1,1)$, $(2,1;1,1)$, $(1,2;1,1)$, and $(1,1;1,1)$, corresponding to the four combinations of genes transmitted by the heterozygous (dihybrid) daughter [or $(4,5;2,3)$, $(4,3;2,3)$, $(2,5;2,3)$, and $(2,3;2,3)$]; and according to the idea of free assortment each of these four combinations should appear with the same relative frequency: $\frac{1}{4}$. But it was observed that the combined frequency of the two types $(1,1;1,1)$ and $(2,2;1,1)$ was larger than that of the types $(2,1;1,1)$ and $(1,2;1,1)$. "The characters that went in together have come out together in a much higher percentage than expected from Mendel's second law, viz. the law of independent assortment" [11]. Morgan, in his theory of the gene called this "tendency" *linkage*. The idea is that the two genes 2,2 and 1,1 which have been together in the maternal individual tend to stay together and that nature has to make an effort to produce a so-called *crossing-over*, i.e. a separation of the genes "that came in together,"—such that a female of type $(2,2;1,1)$ may transmit the group 1,2 or the group 2,1. In other words, *the idea of linkage implies an influence of the grandparents*.

According to observation the percentage of crossing over varies from 0 to 50 per cent, i.e. from *complete linkage* to *free assortment*. It will appear however that in principle crossover-values greater than 50 per cent cannot be excluded. It was also observed that the percentage of individuals of type $(1,1;1,1)$ equals very nearly that of individuals of type $(2,2;1,1)$, as we would expect. In the same way the percentages of types $(2,1;1,1)$ and $(1,2;1,1)$ are nearly equal, their sum yielding the *crossover-ratio*. Hence the four probabilities correspond-

ing to the formation of the four types (1,1; 1,1), (2,2; 1,1), (2,1; 1,1), and (1,2; 1,1) are assumed to be $(1-c)/2$, $(1-c)/2$, $c/2$, and $c/2$. It is important to notice that these are at the same time the probabilities that the female of type (2,2; 1,1) (which was mated to the homozygous (1,1; 1,1), transmits the groups 1,1 or 2,2 or 2,1 or 1,2 respectively.

In the general case of m characters there are $\binom{m}{2} = \frac{m(m-1)}{2}$ crossover probabilities. In this case Morgan assumes *linkage-groups*, each group consisting of m_i elements with $\sum_i m_i = m$, such that "there is linkage between the elements of each group but that the members belonging to different linkage groups assort independently, in accordance with Mendel's second law." This idea will be reconsidered in Section 4.

If we now wish to solve our first basic problem, i.e. to derive the distribution of genotypes in any later generation from an initial distribution of genotypes, then the concept of crossover probabilities does not suffice. The complex possibilities which arise if Mendel's second law is no longer accepted as universally valid cannot be adequately described in terms of crossover probabilities. Or, more exactly: It will be seen that if $m \geq 4$ the crossover probabilities are no longer sufficient, whereas for $m = 2$ and $m = 3$ this concept is general enough. For the complete description of the hereditary mechanism in the general case a so-called *linkage distribution*, i.d., is needed which involves 2^m probabilities with sum equal to one. Let us define this distribution.

Consider an individual of type $(x_1, \dots, x_m; y_1, \dots, y_m) = (x; y)$, where the x are the maternal genes, the genes contributed by the mother of the individual, and the y the paternal genes. Denote by S the set of the m numbers 1, 2, \dots , m , by A any subset of S , and by A' the complementary subset $A' = S - A$. Denote by $l(A)$ the probability that the individual $(x; y)$ transmits the paternal genes belonging to A and the maternal genes belonging to A' . There are $1 + m + \binom{m}{2} + \dots + 1 = 2^m$ such subsets A and accordingly 2^m probabilities $l(A)$ where

$$(8) \quad \sum_{(A)} l(A) = 1.$$

In accordance with the previously reported observations and with our assumption of equal conditions for both sexes one must assume that

$$(8') \quad l(A) = l(A').$$

The conditions (8) and (8') reduce the number of freely disposable values of the l -distribution to $(2^{m-1} - 1)$. The l -distribution is a so-called *m-dimensional* or *m-variate* alternative which could also and occasionally will be denoted by $l(\epsilon_1, \epsilon_2, \dots, \epsilon_m)$ where $\epsilon_i = 0$ or 1. Thus e.g. $l(1, 1, 1, 0, 0, 1)$ is the probability that the genes $y_1, y_2, y_3, x_4, x_5, y_6$, are the genes contained in the germ cell of the individual $(x; y)$. Here the set A consists of the numbers 1, 2, 3, 6, and A' of 4, 5.

Analogous to the statements (a), (b), (c) of §2 and (a'), (b'), (c') of the present section we may now formulate the *principles of Mendel's theory of heredity under the assumption of a possible linkage of the genes*:

(a'') With respect to m Mendelian characters an individual is characterized by two sets of numbers each consisting of m numbers, viz. x_1, \dots, x_m and y_1, \dots, y_m , where $\begin{smallmatrix} x_i \\ y_i \end{smallmatrix} = 1, 2, \dots, r$. If the type of an individual is designated by $(x_1, \dots, x_m; y_1, \dots, y_m) \equiv (x; y)$ where x and y denote the maternal and paternal contributions respectively, then $(x; y) \equiv (y; x)$. Hence there are $N_1 = \frac{1}{2}r^m(r^m + 1)$ types of individuals.

(b'') \equiv (b') In the formation of a new individual each parent transmits to the offspring one set of m genes.

(c'') For each parent, the 2^m probabilities of transmitting any one of these 2^m possible sets are given by a linkage distribution $l(A)$ where A is a subset of the set S consisting of the m numbers $1, 2, \dots, m$, and $l(A)$ is the probability that the transmitted set consists of the paternal genes belonging to A and of the maternal genes belonging to $A' = S - A$, and $l(A) \equiv l(A')$.

4. Some properties of the linkage distribution and of the crossover probabilities. In the following we shall need marginal distributions, that is partial sums, of the probabilities within a distribution. In a usual notation:

$$\begin{aligned}
 l_1(x_1) &= \sum_{x_2} \sum_{x_3} \cdots \sum_{x_m} l(x_1, x_2, \dots, x_m), \dots \dots \dots m \text{ distributions} \\
 l_{12}(x_1, x_2) &= \sum_{x_3} \cdots \sum_{x_m} l(x_1, x_2, \dots, x_m), \dots \dots \dots \binom{m}{2} \text{ distributions} \\
 l_{123 \dots m-1}(x_1, x_2, \dots, x_{m-1}) &= \sum_{x_m} l(x_1, x_2, \dots, x_m), \dots \dots m \text{ distributions} \\
 l_{12 \dots m}(x_1, x_2, \dots, x_m) &= l(x_1, x_2, \dots, x_m) \dots \dots \text{the original distribution.}
 \end{aligned}
 \tag{9}$$

These are general formulae for any discontinuous distribution. But if the distribution happens to be an alternative, as the l.d., where x_i takes only two values, any marginal distribution can be completely characterized by two subsets A and A_1 of S where $A \supset A_1$. Denote by $l_A(A_1)$ the sum of all possible linkage probabilities which contain all points of A_1 and no point of $A - A_1$. If, e.g. $m = 8$ and A consists of $1, 3, 5, 6$ and A_1 of $1, 3, 6$ then $l_A(A_1) = l_{136}(1, 1, 0, 1) = \sum_{x_2, x_4, x_7, x_8} l(1, x_2, 1, x_4, 1, 0, x_7, x_8)$. According to the previous notation we have as usual

$$l_A(A_1) = l(A_1), \text{ or } l_{1,2,\dots,m}(x_1, \dots, x_m) = l(x_1, \dots, x_m)$$

and

$$l_o(O) = 1, \text{ if } A = O \text{ is empty.}$$

We will use for the linkage distribution and their marginal distributions the customary notations or these new notations, whichever is more convenient.⁴

As an immediate consequence of our definitions we get the following properties of the l.d.

(i) If (8) holds for any A than

$$(11) \quad l_A(A_1) = l_A(A - A_1).$$

(ii) As a consequence of (8) it follows (with the notation (9)) that

$$(9') \quad l_i(1) = l_i(0) = \frac{1}{2}.$$

(iii) If c_{ij} denotes the c.p. between i and j , then

$$(12) \quad c_{ij} = c_{ji} = l_{ij}(1,0) + l_{ij}(0,1) = 2l_{ij}(1,0) = 2l_{ij}(0,1).$$

(iv) For any three subscripts i, j, k the "triangular" relation holds

$$(13) \quad c_{ij} + c_{jk} \geq c_{ik}$$

and

$$(14) \quad c_{ij} + c_{jk} + c_{ik} \leq 2.$$

To prove this consider the marginal distribution $l_{i,jk}(x_i x_j x_k)$. From (11) and (12) we conclude

$$c_{ij} = 2[l_{i,jk}(100) + l_{i,jk}(010)]$$

$$c_{ik} = 2[l_{i,jk}(100) + l_{i,jk}(001)]$$

$$c_{jk} = 2[l_{i,jk}(010) + l_{i,jk}(001)]$$

$$1 = 2[l_{i,jk}(000) + l_{i,jk}(100) + l_{i,jk}(010) + l_{i,jk}(001)].$$

⁴ It is easy to indicate experiments which should furnish the relative frequencies corresponding to the l.d.: If a homozygous female ($x_1, \dots, x_m; x_1, \dots, x_m$) is mated to a homozygous male ($y_1, \dots, y_m; y_1, \dots, y_m$) where each $x_i \neq y_i$, the resulting offsprings will all be of type ($x_1, \dots, x_m; y_1, \dots, y_m$). If such an offspring is back crossed to ($y_1, \dots, y_m; y_1, \dots, y_m$) there will be 2^m different genotypes of offsprings, viz. ($x_1, x_2, \dots, x_m; y_1, y_2, \dots, y_m$), ($y_1, x_2, \dots, x_m; y_1, y_2, \dots, y_m$), etc. whose frequencies are proportional to the 2^m values of the l.d., viz. to $l(0, 0, \dots, 0)$, $l(1, 0, 0, \dots, 0)$ etc. Such an experiment should give the same results for any two sets of x 's and y 's. (There is, of course, the statistical problem how to determine the "best" values of the l.p. from these observations.) In an analogous way a marginal distribution can be observed: Suppose we wish for $m = 5$, the $l_{123}(e_1, e_2, e_3)$. The offspring of a cross between females ($x_1, x_2, x_3, x_4, x_5; x_1, x_2, x_3, x_4, x_5$) and males ($y_1, y_2, y_3, x_4, x_5; y_1, y_2, y_3, x_4, x_5$) are of type ($x_1, x_2, x_3, x_4, x_5; y_1, y_2, y_3, x_4, x_5$). If they are crossed to ($x_1, \dots, x_5; x_1, \dots, x_5$) there will be eight different types of offsprings proportional to the eight values of $l_{123}(e_1, e_2, e_3)$. In this last setup the y_i should be dominant and in the experiment, described above, the y_i should be recessive in order to be able to distinguish between the phenotypes of the individuals.

Solving these equations with respect to the l -values we get

$$(15) \quad \begin{aligned} l_{i,jk}(100) &= \frac{1}{4} (c_{ij} + c_{ik} - c_{jk}) \\ l_{i,jk}(010) &= \frac{1}{4} (c_{ij} + c_{jk} - c_{ik}) \\ l_{i,jk}(001) &= \frac{1}{4} (c_{ik} + c_{jk} - c_{ij}) \end{aligned}$$

$$(16) \quad l_{i,jk}(000) = \frac{1}{4} (2 - c_{ij} - c_{ik} - c_{jk}).$$

Thence (13) and (14) follow. The condition (14) is of course always fulfilled if $c_{ij} \leq \frac{1}{2}$, but this restriction does not seem to be necessary. From (15) and (16) we deduce:

(v) If $m = 3$, the set of three c.p. c_{12} , c_{13} , c_{23} for which the inequalities (13), (14) hold is equivalent to the l.d. $l(x_1, x_2, x_3)$ for which (8) holds. For $m \geq 4$ the c.p. are no longer equivalent to the l.d. Another necessary condition for the c.p. will be derived in section 8.

Now let us consider and characterize some important particular cases of the l.d.

(i) *Free assortment* (Mendel). In this case all 2^m values of the l.d. are equal and therefore equal to $(\frac{1}{2})^m$.

(ii) *Complete linkage* (reported by Morgan and other authors). In terms of the l.d. this means

$$(17) \quad l(1, 1, \dots, 1, 1) = l(0, 0, \dots, 0, 0) = \frac{1}{2} \text{ or } l_N(S) = \frac{1}{2}.$$

Consequently, all other values of the l.d. are zero. It follows that all c.p. are zero because all $l_{ij}(1, 0)$ are zero. (See also Theorem I, section 7.)

(iii) *Linkage groups* (Morgan). In terms of the l.d. this means that the l.d. resolves into a product of several distributions, e.g.

$$(18) \quad l(x_1, x_2, \dots, x_9) = f(x_1, x_2)g(x_3, x_4, x_5)h(x_6, x_7, x_8, x_9).$$

(There is no loss of generality in assuming that numerically consecutive characters form a linkage group.) As f , g , and h are distributions it follows with notation (9) that "within" the groups:

$$\begin{aligned} c_{12} = 2f(10), \quad c_{34} = 2g_{34}(10), \dots, \quad c_{45} = 2g_{45}(10), \\ c_{67} = 2h_{67}(10), \dots, \quad c_{89} = 2h_{89}(10) \end{aligned}$$

these crossover values are quite arbitrary. On the other hand we have because of (9')

$$\begin{aligned} f_i(1) = f_i(0) = g_j(1) = g_j(0) = h_k(1) = h_k(0) = \frac{1}{2}, \\ (i = 1, 2; j = 1, 2, 3; k = 1, \dots, 4) \end{aligned}$$

Hence for the c.p. "among" the groups

$$c_{13} = 2 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}, \text{ etc. Hence } c_{13} = c_{14} = c_{23} = \dots = c_{49} = \frac{1}{2}$$

in exact accordance with Morgan's idea of linkage groups. If each group consists of only one element: $l(x_1, x_2, \dots, x_m) = f(x_1)g(x_2) \dots k(x_m)$ it follows

that $f(x_1) = g(x_2) = \dots = k(x_m) = \frac{1}{2}$ for $x_i = 0, 1$, hence $l(x_1, \dots, x_m) = (\frac{1}{2})^m$ for all combinations of the arguments and we have again free assortment.

(iv) *Groups of completely linked characters.* Combining and generalizing the ideas of (ii) and (iii) we may speak of i groups of completely linked characters if *within such a group no crossover takes place*. Then the m_i characters in each group act as one character. An example will suffice. Suppose $m = 9$ and three such groups, consisting of the characters 1, 2, and 3, 4, 5, and 6, 7, 8, 9 respectively. Assume that

$$l(11, 111, 1111) = l(00, 000, 0000) = a, \quad l(00, 111, 1111) = l(11, 000, 0000) = C_1 \\ l(11, 000, 1111) = l(00, 111, 0000) = C_2, \quad l(11, 111, 0000) = l(00, 00, 1111) = C_3$$

where these four numbers are $\neq 0$ and with sum $\frac{1}{2}$; hence all other probabilities are zero. It follows that the c.p. "within" the groups are all zero: $c_{12} = c_{24} = \dots = c_{87} = c_{89} = 0$, but the "among" c.p. are different from zero, e.g. $c_{13} = c_{14} = c_{15} = c_{23} = c_{24} = c_{25} = 2C_1 + 2C_2$ and, with an obvious notation: $c_{I,II} = 2(C_1 + C_2)$, $c_{I,III} = 2(C_1 + C_3)$, $c_{II,III} = 2(C_2 + C_3)$.

A particular case (also a particular case of (iii)) arises if the l.d. resolves into a product of some distributions such that there is complete linkage in each of these. The "within" crossovers are then again zero but all the c.p. "among" the groups equal $\frac{1}{2}$.

5. The case $m = 2$. It will be easier for the reader if this case, though it has been investigated before by several authors [16], [7], [14], will be presented by means of explicit computations before attempting the general one where m and r are arbitrary.

If $m = r = 2$, the number of types $(x_1, x_2; y_1, y_2)$ equals ten. The l.d. is completely determined by the c.p. $c_{12} = c$ and v.v., because $l(10) = l(01) = c/2$, $l(00) = l(11) = (1 - c)/2$. Now let $p^{(n)}(x_1, x_2)$ be the probability that in the n th generation a male (or female) individual *transmits the genes* x_1, x_2 ; and denote by $p_1^{(n)}(x_1)$ and $p_2^{(n)}(x_2)$ the respective marginal distributions. The formula corresponding to (4) then becomes

$$(19) \quad p^{(n)}(1, 1) = v^{(n)}(1, 1; 1, 1) + \frac{1}{2}v^{(n)}(1, 1; 1, 2) + \frac{1}{2}v^{(n)}(1, 1; 2, 1) \\ + \frac{1 - c}{2}v^{(n)}(1, 1; 2, 2) + \frac{c}{2}v^{(n)}(1, 2; 2, 1),$$

and three analogous formulae. To understand this, consider e.g. the last term of (19); it is the probability that an individual be of type $(1, 2; 2, 1)$ or $(2, 1; 1, 2)$ and transmits the set $(1, 1)$. By (19) $p^{(n)}(x_1, x_2)$ is deduced from the given distribution $v^{(n)}$ of genotypes.

If, as before, x and y are written for x_1, x_2 and y_1, y_2 it is to be understood that $x = y$ means $x_1 = y_1$ and $x_2 = y_2$. The relation corresponding to (5') takes then the form

$$(20) \quad v^{(n+1)}(x; y) = p^{(n)}(x)p^{(n)}(x) \quad \text{if } x = y \\ = 2p^{(n)}(x)p^{(n)}(y) \quad \text{if } x \neq y.$$

Applying (19) to the $(n + 1)$ st generation and using (20) we get the recurrence formula

$$(21) \quad p^{(n+1)}(1,1) = [p^{(n)}(1,1)]^2 + p^{(n)}(1,1)p^{(n)}(1,2) + p^{(n)}(1,1)p^{(n)}(2,1) \\ + (1-c)p^{(n)}(1,1)p^{(n)}(2,2) + cp^{(n)}(1,2)p^{(n)}(2,1).$$

Here the right side can be rewritten so as to give

$$(22) \quad p^{(n+1)}(1,1) = (1-c)p^{(n)}(1,1) + cp_1^{(n)}(1)p_2^{(n)}(1)$$

and three analogous formulae. Because of (7):

$$(22') \quad p^{(n+1)}(x_1, x_2) = (1-c)p^{(n)}(x_1, x_2) + cp_1^{(n)}(x_1)p_2^{(n)}(x_2).$$

From this recurrence formula, which has the particularly simple property that the second term on the right side is independent of n , it is easy to derive step by step:

$$(23) \quad p^{(n)}(x_1, x_2) = (1-c)^n p^{(0)}(x_1, x_2) + [1 - (1-c)^n] p_1^{(0)}(x_1) p_2^{(0)}(x_2).$$

Hence, if $c \neq 0$:

$$(24) \quad \lim_{n \rightarrow \infty} p^{(n)}(x_1, x_2) = p_1^{(0)}(x_1) p_2^{(0)}(x_2).$$

The preceding results were obtained by Robbins and Jennings. We will formulate a theorem after having studied the general case of arbitrary m and r .⁷

6. The general recurrence formula. Considering random mating and assuming general linkage, we now wish to find the relations which correspond to the formulae (19)–(22) in the case of m r -valued characters. It will turn out, that, by using the l.d., the following proof of the general case becomes surprisingly simple compared with older investigations of the particular case of free assortment, the values of the l.d. acting somehow as natural "separators" for certain groups of terms.

Denote by $w^{(n)}(x_1, \dots, x_m; y_1, \dots, y_m) \equiv w^{(n)}(x; y)$ the probability of a genotype whose maternal genes are the x and whose paternal genes the y . Then from (a''):

$$(25) \quad w^{(n)}(x; y) = w^{(n)}(y; x).$$

Writing $x = y$ if and only if $x_i = y_i$, ($i = 1, \dots, m$) we put just as in (2)

$$(25') \quad v^{(n)}(x; y) = w^{(n)}(x; x), \quad \text{if } x = y \\ = w^{(n)}(x; y) + w^{(n)}(y; x) = 2w^{(n)}(x; y), \quad \text{if } x \neq y.$$

⁷ A suggestive remark, repeatedly made by Professor S. Wright states that (assuming random mating) there can be no equilibrium until all of the factors are combined at random. This is indeed a necessary condition for stability.

There are r^{2m} w -values and $\frac{1}{2}r^m(r^m + 1)$ v -values in each generation the respective sums being always equal to one. Denote by $p^{(n)}(x_1, \dots, x_m)$ the probability that a male (female) individual of the n th generation transmits the genes x , and by

$$p_1^{(n)}(x_1), p_{12}^{(n)}(x_1, x_2), \dots, p_{12\dots m}^{(n)}(x_1, \dots, x_m) = p^{(n)}(x_1, \dots, x_m)$$

the corresponding marginal distributions, defined as usual (see (9)). Sometimes it will be convenient to denote such a marginal distribution by $p_A(z_A) = p_A(z)$ where $A \subset S$, and $p_A(z)$ is the sum of all $p(x)$ such that $x_i = z_i$ for all $i \in A$. Following convention the subscript will be omitted if $A = S$; hence $p_S(z) = p(z)$ and if A is empty, $A = \emptyset$, the corresponding $p_\emptyset(z) = 1$.

To simplify the writing $p(x)$, $v(x; y)$, etc. will be written instead of $p^{(n)}(x)$, $v^{(n)}(x; y)$, etc. and $p'(x)$, $v'(x; y)$, etc. for $p^{(n+1)}(x)$, etc. Finally, remember that $l(A)$ is the probability that the paternal genes of A and the maternal genes of $A' = S - A$ will be transmitted and accordingly $l_1(A_1)$ is the (marginal) probability that the paternal genes of A_1 and the maternal genes of $A - A_1$ will be transmitted. ($S \supset A \supset A_1$).

Let us derive $p'(z)$ from $p(z)$. From the meaning of the different distributions we gather that

$$(26) \quad p(z) = \Sigma l(A)w(x; y)$$

where A is an arbitrary subset of S and x and y such that

$$(a) \quad \begin{aligned} y_i &= z_i & \text{for } i \in A \\ x_i &= z_i & \text{" } i \in A'. \end{aligned}$$

In fact, the set z will be transmitted if and only if an individual possesses these genes and also transmits them; now consider any $l(A)$, i.e. the probability to transmit the paternal genes of A ; this probability is to be multiplied by all possible w -probabilities which contain as arguments the paternal genes of A and the maternal genes of A' , as stated in (a). Now let us write (26) also for the $(n + 1)$ st generation:

$$(26') \quad p'(z) = \Sigma l(A)w'(x; y).$$

Next we have, just as always, [see (5), (20)]

$$(27) \quad w'(x; y) = w'(y; x) = p(x)p(y).$$

Hence from (26') and (27) follows

$$(28) \quad p'(z) = \Sigma l(A)p(x)p(y)$$

with the condition of summation given by (a).

The right side of (28) contains $(2r)^m$ terms. Now we will write it in two different ways by collecting its terms under two different aspects: (i) arranged according to the marginal values of the l -distribution (ii) arranged according to the marginal values of the p -distribution. Let us begin with (i).

The genes z_1, z_2, \dots, z_m can be transmitted only by individuals which possess each z_i either before or after the semicolon or both; (either from the mother or from the father or from both parents). Hence, if A_1 and A_2 are two disjoint subsets of S , the type of such an individual is such that

$$\begin{aligned} (b) \quad & x_i \neq z_i, \quad y_i = z_i \quad \text{for all } i \in A_1 \\ & x_j = z_j, \quad y_j \neq z_j \quad \text{" " } j \in A_2 \\ & x_k = y_k = z_k \quad \text{" " } k \in S - A_1 - A_2. \end{aligned}$$

Hence the paternal genes of A_1 and the maternal genes of A_2 must be transmitted and for the remaining genes either choice is admissible. Consequently, each $w'(x; y)$ in (26')—or, what is the same, each $p(x)p(y)$ in (28)—is multiplied by the probability that the paternal genes of A_1 and the maternal genes of A_2 are transmitted. Now writing $A_1 + A_2 = A$ this last probability is exactly the marginal probability $l_A(A_1) = l_A(A_2)$. Thence

$$(29) \quad p'(z) = \sum p(x)p(y)l_A(A_1)$$

where the sum is extended over all pairs x, y defined by (b). This is a first recurrence formula. If in (29) $w'(x; y)$ is written instead of $p(x)p(y)$ and then all accents are omitted we get

$$(30) \quad p(z) = \sum w(x; y)l_A(A_1)$$

with the summation according to (b). This formula is necessary in order to derive $p(z)$ from the given distribution $w(x; y)$ of genotypes. It corresponds to (19).

Now let us collect the terms of (28) in the second way. Let us determine the factor of any $l(A)$ in (28), e.g. of $l(1, 1, 0, 0, 0)$ (where $m = 5$ and A the subset 1, 2). Any factor of $l(1, 1, 0, 0, 0)$ must be of the form $p(z_1, z_2, \cdot, \cdot, \cdot) p(\cdot, \cdot, z_3, z_4, z_5)$ where all possible values of the variables must be written on the empty places marked by points, and the sum of all these products is to be taken. Now, as in each of the two p 's on each of the free places all numbers between 1 and r have to be used, the sum of all these products resolves into the product of the respective sums of the p 's. In such a sum each term, on the places belonging to A contains the same fixed values z_A and on the other places any possible value combination; hence such a sum is precisely the marginal probability $p_A(z_A) = p_A(z)$ and the same holds for the other sum of the p 's and for $A' = S - A$. Thus we get the second, even more important recurrence formula

$$(31) \quad p'(z) = \sum_{(A)} l(A)p_A(z)p_{A'}(z)$$

where the sum is over all subsets A of S . This formula corresponds to (22) and the limit theorem which will be proved in the next section is an almost immediate consequence of (31). It is worth noticing that the derivations of

(31) and (29) from (28) are completely independent of each other and that only (31) is needed for the limit theorem

From (29) and (31) the interesting identity follows

$$(32) \quad \sum_{(A)} l(A) p_A(z) p_{A'}(z) = \sum p(x) p(y) l_A(A_1)$$

which somehow reminds us of a general Abel-transformation.

Let us summarize: (i) From a given distribution of genotypes $w^{(n)}$ (or $v^{(n)}$) the $p^{(n)}$ are derived by (30). (ii) From these $p^{(n)}$ the $w^{(n+1)}$ follow by (27) (or $v^{(n+1)}$ by (25') and (27')). (iii) Instead of step (ii), from $p^{(n)}$ the consecutive $p^{(n+1)}$, $p^{(n+2)}$, ..., $p^{(n+r)}$ may be derived directly by means of (31). Finally, if desired, $w^{(n+r+1)}$ follows by (27).

As an illustration of these formulae let us write (31) for $m = 3, 4, 5$:

$$(31') \quad \begin{aligned} p'(x_1, x_2, x_3) &= 2[l(000)p(x_1, x_2, x_3) + l(100)p_1^{(0)}(x_1)p_{23}(x_2, x_3) \\ &+ l(010)p_2^{(0)}(x_2)p_{13}(x_1, x_3) \\ &+ l(001)p_3^{(0)}(x_3)p_{12}(x_1, x_2)] \end{aligned}$$

$$(31'') \quad \begin{aligned} p'(x_1, x_2, x_3, x_4) &= 2[l(0000)p(x_1, x_2, x_3, x_4) \\ &+ l(1000)p_1^{(0)}(x_1)p_{234}(x_2, x_3, x_4) + \dots \\ &+ l(1100)p_{12}(x_1, x_2)p_{34}(x_3, x_4) \\ &+ l(1010)p_{13}(x_1, x_3)p_{24}(x_2, x_4) \\ &+ l(1001)p_{14}(x_1, x_4)p_{23}(x_2, x_3)] \end{aligned}$$

$$(31''') \quad \begin{aligned} p'(x_1, x_2, x_3, x_4, x_5) &= 2[l(00000)p(x_1, \dots, x_5) \\ &+ l(10000)p_1^{(0)}(x_1)p_{2345}(x_2, x_3, x_4, x_5) + \dots \\ &+ l(11000)p_{12}(x_1, x_2)p_{345}(x_3, x_4, x_5) + \dots]. \end{aligned}$$

In the last formula the last group contains ten terms. As an illustration of (30) we write e.g. for $m = 3$, $r = 2$, with $p^{(n)} = p$ and $v^{(n)} = v$:

$$(30') \quad \begin{aligned} p(x_1, x_2, x_3) &\equiv v(x_1, x_2, x_3; x_1, x_2, x_3) + \frac{1}{2}[v(x_1, x_2, x_3; y_1, x_2, x_3) \\ &+ v(x_1, x_2, x_3; x_1, y_2, x_3) + \dots] \\ &+ [l_{12}(00)v(x_1, x_2, x_3; y_1, y_2, x_3) \\ &+ l_{13}(00)v(x_1, x_2, x_3; y_1, x_2, y_3) + \dots] \\ &+ l(000)v(x_1, x_2, x_3; y_1, y_2, y_3) \\ &+ [l(100)v(y_1, x_2, x_3; x_1, y_2, y_3) \\ &+ l(010)v(x_1, y_2, x_3; y_1, x_2, y_3) + \dots]. \end{aligned}$$

7. Limit theorems. In order to find $\lim_{n \rightarrow \infty} p^{(n)}(x_1, \dots, x_m)$ we write the recurrence formula (31) in the form

$$(33) \quad p^{(n+1)}(x) = 2l(00 \dots 0)p^{(n)}(x) = \sum_{(A)} l(A)p_A^{(n)}(z)p_A^{(n)}(z).$$

Here $\sum_{(A)}$ means a sum over all subsets A of S which are neither void nor equal to S . If we write $q_m^{(n)}$ for the right side of (33) and $p^{(n)}(x) = p_m^{(n)}$, $2l(0, \dots, 0) = \alpha_m$ the last equation takes the form

$$(34) \quad p_m^{(n+1)} = \alpha_m p_m^{(n)} = q_m^{(n)}.$$

Consider first the case $\alpha_m = 1$, or $l(0, \dots, 0) = l(1, \dots, 1) = \frac{1}{2}$, i.e. *complete linkage*, as defined in section 3. In this case all $l(A)$ -values on the right side of (33) are zero, hence $q_m^{(n)} = 0$ and

$$(35) \quad p_m^{(n+1)} = p_m^{(n)} \quad (n = 0, 1, 2, \dots).$$

This is exactly the same result as (7): All $p_m^{(n)}$ are equal to $p_m^{(0)}$ and because of (27) also

$$(36) \quad w^{(n)}(x; y) = w^{(1)}(x; y) \quad \text{or} \quad v^{(n)}(x; y) = v^{(1)}(x; y) \quad (n = 1, 2, \dots).$$

In fact, if the characters are completely linked, they act as one character. Hence we have

THEOREM I. *If the m Mendelian characters are completely linked, the distribution of genotypes reaches the stationary state in the first filial generation.*

Now consider (34) in the general case where $0 \leq \alpha_m < 1$. Then the following lemma will be used: *If in a recurrence formula of the form (34), $\alpha_m < 1$ and $\lim_{n \rightarrow \infty} q_m^{(n)} = q_m$ exists, then $\lim_{n \rightarrow \infty} p_m^{(n)} = p_m = q_m/(1 - \alpha_m)$.* This can be proved directly in various simple ways. It may also be regarded as a consequence of well-known general convergence theorems. See also [15].

In order to apply the lemma let us first notice that q_2 exists. In fact, $p^{(n+1)}(x_1, x_2) = 2l(00)p^{(n)}(x_1, x_2) = 2l(01)p_1^{(0)}(x_1)p_2^{(0)}(x_2)$ and as the right side is independent of n , q_2 certainly exists. Hence, it follows from the lemma that p_2 exists. For $m = 3$ the recurrence formula (31') shows that $q_3^{(n)}$ contains no marginal distribution of p of an order higher than two; therefore each of the terms of $q_3^{(n)}$ approaches a limit, hence $q_3 = \lim_{n \rightarrow \infty} q_3^{(n)}$ exists, and consequently,

because of the lemma, p_3 exists. We may continue in this way because in (33) all marginal distributions of p on the right side are of an order $\leq m - 1$. Hence for every m the $q_m^{(n)}$ approaches a limit and consequently the $\lim_{n \rightarrow \infty} p_m^{(n)}$ exists.

Finally, in order to find p_m we notice that $q_2 = (1 - \alpha_2)p_1^{(0)}(x_1)p_2^{(0)}(x_2)$, hence $p_2 = p_1^{(0)}(x_1)p_2^{(0)}(x_2)$. Then, assuming that $p_{m-1} = p_1^{(0)}(x_1) \dots p_{m-1}^{(0)}(x_{m-1})$ we see from (33), using (8), that $q_m = (1 - \alpha_m)p_1^{(0)}(x_1) \dots p_m^{(0)}(x_m)$. (See also (31') (31''), (31''').) Thence

$$(37) \quad \lim_{n \rightarrow \infty} p^{(n)}(x_1, x_2, \dots, x_m) = p_1^{(0)}(x_1)p_2^{(0)}(x_2) \dots p_m^{(0)}(x_m).$$

The last formula contains the limit theorem we wished to prove. It can be stated as follows:

THEOREM II. *If m characters are arbitrarily linked, with the one exception of "complete linkage", the distribution of transmitted genes $p^{(n)}(x_1, \dots, x_m)$ "converges towards independence." The limit distribution is the product of the m marginal distributions of the first order $p_i^{(0)}(x_i)$, which are derived from $p^{(0)}(x_1, \dots, x_m)$, the distribution of gametes in the initial generation.*

If, however, the initial distribution $p^{(0)}(x_1, \dots, x_m)$ shows particular features, the stationary state may be reached already for a finite value of n . This happens with $n = 0$ and for every l.d. if $p^{(0)}(x_1, \dots, x_m) = p_1^{(0)}(x_1) \cdots p_m^{(0)}(x_m)$. In other particular cases it may happen under particular assumptions for the l.d.

Let us express the general result also in terms of the distribution of genotypes. It follows from (37) and (27) that

$$\begin{aligned} \lim_{n \rightarrow \infty} w^{(n+1)}(x; y) &= \lim_{n \rightarrow \infty} p^{(n)}(x) p^{(n)}(y) \\ &= p_1^{(0)}(x_1) \cdots p_m^{(0)}(x_m) p_1^{(0)}(y_1) \cdots p_m^{(0)}(y_m) = \prod_{i=1}^m [p_i^{(0)}(x_i) p_i^{(0)}(y_i)]. \end{aligned}$$

Now consider a product like $p_i^{(0)}(x_i) p_i^{(0)}(y_i)$. By definition of $p_i^{(0)}(x_i)$ and applying (27) we find

$$\begin{aligned} p_1^{(0)}(x_1) p_1^{(0)}(y_1) &= \sum_{x_1} \cdots \sum_{x_m} p^{(0)}(x_1, \dots, x_m) \sum_{y_1} \cdots \sum_{y_m} p^{(0)}(y_1, \dots, y_m) \\ &= \sum_{x_1, \dots, x_m} \sum_{y_1, \dots, y_m} p^{(0)}(x) p^{(0)}(y) \end{aligned}$$

Introducing then in a natural way the marginal distribution:

$$(38) \quad w_i^{(n)}(x_i; y_i) = \sum_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m} \sum_{y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_m} w^{(n)}(x_1, \dots, x_m; y_1, \dots, y_m)$$

it is seen that

$$(39) \quad p_i^{(0)}(x_i) p_i^{(0)}(y_i) = w_i^{(1)}(x_i; y_i).$$

Thence the result

$$(40) \quad \lim_{n \rightarrow \infty} w^{(n)}(x_1, \dots, x_m; y_1, \dots, y_m) = w_1^{(1)}(x_1; y_1) \cdots w_m^{(1)}(x_m; y_m)$$

which may be stated as follows:

THEOREM III. *In case of m arbitrarily linked Mendelian characters the distribution of the genotypes in the n th generation, $w^{(n)}(x_1, \dots, x_m; y_1, \dots, y_m)$, "approaches independence" as $n \rightarrow \infty$. The limit distribution is the product of the m marginal distributions $w_i^{(1)}(x_i; y_i)$ of the i th character ($i = 1, \dots, m$) in the first filial generation.*

This theorem, which may be regarded as a corollary to THEOREM II, holds for any type of linkage, except "complete linkage" as defined in (17) where (36) is valid.

8. Solution of the recurrence equations (31). Formula (31) expresses $p^{(n)}(x_1, \dots, x_m)$ in terms of $p^{(n-1)}(x_1, \dots, x_m)$ (and all marginal distributions of $p^{(n-1)}$) and of the l.d. It seems desirable to try to express $p^{(n)}(x)$ in terms of $p^{(0)}(x)$. Now (31) is not a single equation but rather a complex system of difference equations with constant coefficients because for each marginal distribution of order $i < m$ the respective recurrence formula (31) of order i has to be used. (Or, if it is preferred to consider the marginal distributions as sums of p -values of order m , then all these p -values appear simultaneously and there is again a complicated system of difference equations.) In this situation it is not to be expected that the integration will yield simple explicit formulae, particularly as long as the l.d. is left arbitrary. However, the construction of the following formulae is clear. They reduce to simpler expressions in particular cases.

Let us use a method of indeterminate coefficients. To simplify the writing denote $p^{(0)}(x_1, \dots, x_m)$ and its marginal distributions $p_{i_1}^{(0)}(x_1)$, $p_{i_1 i_2}^{(0)}(x_1, x_2)$, etc. by $p_{12}, \dots, p_i, p_{ij}$, etc. From genetical as well as mathematical considerations we gather the general form of $p_{12 \dots m}^{(n)}$ in terms of $p_{12 \dots m}$ and its marginal distributions; that this is indeed the general form will be verified by our very computations. Consider the set S consisting of the m numbers $1, 2, \dots, m$ and divide S in every possible way in two disjoint parts A_1 and A_2 , none of them being empty, so that $A_1 + A_2 = S$, then divide S in every possible way into three disjoint parts so that $A_1 + A_2 + A_3 = S$, and finally S is divided into m disjoint parts each consisting of one single element. Denoting the unknown coefficients in a corresponding way by $\alpha_s^{(n)}$, $\alpha_{A_1 A_2}^{(n)}$, $\alpha_{A_1 A_2 A_3}^{(n)}$, etc. and writing $p_i^{(n)}$ and p_s for $p_{12 \dots m}^{(n)}$ and $p_{12 \dots m}^{(0)}$ the general form of p_s will be

$$(41) \quad p_s^{(n)} = \alpha_s^{(n)} p_s + \sum_{(A_1)} \alpha_{A_1}^{(n)} p_{A_1} p_{A_2} + \sum_{(A_1 A_2)} \alpha_{A_1 A_2}^{(n)} p_{A_1} p_{A_2} p_{A_3} \\ + \dots + \alpha_{1,2,3,\dots,m}^{(n)} p_1 p_2 p_3 \dots p_m.$$

This holds for every m . We get e.g. for $m = 4$

$$(41') \quad p_{1234}^{(n)} = \alpha_{1234}^{(n)} p_{1234} + (\alpha_{1,234}^{(n)} p_1 p_{234} + \alpha_{2,134}^{(n)} p_2 p_{134} + \dots) \\ + (\alpha_{12,34}^{(n)} p_{12} p_{34} + \dots) + (\alpha_{12,3,4}^{(n)} p_{12} p_3 p_4 + \dots) + \alpha_{1,2,3,4}^{(n)} p_1 p_2 p_3 p_4.$$

For $m = 6$, e.g., there are eleven different types: One term $\alpha_1^{(n)} p_1$; then 6 terms of the form $\alpha_{1,2 \dots 5}^{(n)} p_{12 \dots 5}$; 15 terms like $\alpha_{12,3456}^{(n)} p_{12} p_{3456}$; 10 terms like $\alpha_{123,456}^{(n)} p_{123} p_{456}$; 15 terms like $\alpha_{1,2,3456}^{(n)} p_1 p_{23456}$; 60 terms like $\alpha_{1,2,3,456}^{(n)} p_1 p_2 p_{3456}$; 15 terms like $\alpha_{12,34,56}^{(n)} p_{12} p_{34} p_{56}$; 20 terms as $\alpha_{1,2,3,4,56}^{(n)} p_1 p_2 p_3 p_{456}$; 15 terms as $\alpha_{12,34,5,6}^{(n)} p_{12} p_{34} p_5 p_6$; 15 terms as $\alpha_{1,2,3,4,5,6}^{(n)} p_1 p_2 p_3 p_4 p_5 p_6$; and one final term

In (41) the $\alpha_s^{(n)}$ are unknown constants depending on n and on the l.d. In order to find them consider (31) and write for the values of the l.d. x_A^m instead of $2l(A)$ (no confusion is possible because no marginal distribution of the l.d. occurs in (31)). With this notation (31'') e.g. reads:

$$(31'') \quad p_{1234}^{(n+1)} = v_0^4 p_{1234}^{(n)} + (v_1^4 p_1 p_{234}^{(n)} + \dots) + (v_{12}^4 p_{12}^{(n)} p_{34}^{(n)} + \dots).$$

If there is no ambiguity the upper m in v_μ^m may even be omitted. Now assume the equations (41) to be written for $\mu = 2, \mu = 3, \dots, \mu = m$. Introduce into the left side of (31) the expression (41) for $p_\mu^{(n-1)}$ and in the same way replace on the right side of (31) all $p_i^{(n)}, p_j^{(n)}, \dots, p_{i_2}^{(n)}$ by their respective expressions (41). In this way an equality is obtained from which recurrence formulae for the unknown coefficients may be deduced by collecting all groups of terms which contain the same products of p 's.

If this is carried out, e.g. for $m = 4$, the recurrence formulae are

$$\begin{aligned}
 \alpha_{1234}^{(n+1)} &= v_0 \alpha_{1234}^{(n)} \\
 \alpha_{123,4}^{(n+1)} &= v_0 \alpha_{123,4}^{(n)} + v_4 \alpha_{123}^{(n)} \\
 \alpha_{12,34}^{(n+1)} &= v_0 \alpha_{12,34}^{(n)} + v_{12} \alpha_{12}^{(n)} \alpha_{34}^{(n)} \quad \text{etc.} \\
 \alpha_{12,3,4}^{(n+1)} &= v_0 \alpha_{12,3,4}^{(n)} + v_{12} \alpha_{12}^{(n)} \alpha_{3,4}^{(n)} + v_3 \alpha_{12,4}^{(n)} + v_4 \alpha_{12,3}^{(n)} \\
 \alpha_{1,2,3,4}^{(n+1)} &= v_0 \alpha_{1,2,3,4}^{(n)} + v_{12} \alpha_{2,3,4}^{(n)} + \dots + v_{12} \alpha_{1,2}^{(n)} \alpha_{3,4}^{(n)} + \dots
 \end{aligned}
 \tag{42}$$

In general, i.e. for any m , these recurrence formulae are of a clear structure the first one being particularly simple, namely

$$\alpha_s^{(n+1)} = v_0 \alpha_s^{(n)}.$$

It can be solved immediately and gives

$$\alpha_s^{(n)} = v_0^n.$$

The other recurrence formulae are all of the form

$$x_{n+1} = v_0 x_n + f(n) \text{ with } x_0 = 0,$$

where $f(n)$ is a given function of n whose general form is still to be investigated. The solution of (44) is

$$x_n = \sum_{\nu=0}^{n-1} f(\nu) v_0^{n-1-\nu}.$$

With the notations used in (41) the equation (44) may be written:

$$\alpha_{A_1, A_2, \dots, A_\mu}^{(n+1)} = v_0 \alpha_{A_1, A_2, \dots, A_\mu}^{(n)} + A_{A_1, A_2, \dots, A_\mu}^{(n)}.$$

We have to determine $A_{A_1, A_2, \dots, A_\mu}$. For reasons of symmetry and homogeneity let us introduce constants $\alpha_1^{(n)} = \alpha_2^{(n)} = \dots = \alpha_m^{(n)} = 1$. With that notation e.g. the last term in the second line in (42) reads $v_4 \alpha_{123}^{(n)} \alpha_4^{(n)}$ or the third term to the right in the fourth line of (42): $v_3 \alpha_{12,4}^{(n)} \alpha_3^{(n)}$ etc.

The construction of $A_{A_1, A_2, \dots, A_\mu}^{(n)}$ may then be described as follows: Each $A_{A_1, A_2, \dots, A_\mu}^{(n)}$ is a sum of $2^{\mu-1} - 1$ terms, each term being a product of one v -value and two α 's. The set consisting of the μ elements A_1, A_2, \dots, A_μ is to be divided in all possible ways into two non-empty, disjoint, complementary parts which form the subscripts of the two α 's in question; the subscript of v is equal to the subscript of either of these two α -values; it makes no difference which,

because of the specific symmetry (8') of the l.d.; it should be noted that in the subscripts of v no comma occurs. As an example let us write $A_{1234,567,8}^{(n)}$ for $m = 8$. We get: $A_{1234,567,8}^{(n)} = v_8 \alpha_8^{(n)} \alpha_{1234,567}^{(n)} + v_{567} \alpha_6^{(n)} \alpha_{1234,8}^{(n)} + v_{1234} \alpha_{12}^{(n)} \alpha_{567,8}^{(n)}$. (Or if we wish $A_{12,34,5,6}^{(n)}$ for $m = 6$: $A_{12,34,5,6}^{(n)} = v_6 \alpha_6^{(n)} \alpha_{12,34,5}^{(n)} + v_5 \alpha_5^{(n)} \alpha_{12,34,6}^{(n)} + v_{12} \alpha_{12}^{(n)} \alpha_{34,5,6}^{(n)} + v_{34} \alpha_{34}^{(n)} \alpha_{12,5,6}^{(n)} + v_{56} \alpha_{5,6}^{(n)} \alpha_{12,34}^{(n)} + v_{125} \alpha_{12,5}^{(n)} \alpha_{34,6}^{(n)} + v_{126} \alpha_{12,6}^{(n)} \alpha_{34,5}^{(n)}$.

Hence, in principle our "integration" problem, where n is the variable, is completely solved: First $p_s^{(n)}$ is given by (41). Then, in order to find any $\alpha_{A_1, A_2, \dots, A_\mu}^{(n)}$, we first determine the corresponding A_{1, A_2, \dots, A_μ} by the rule just explained and illustrated, and then it follows from (44') that

$$(44''') \quad \alpha_{A_1, A_2, \dots, A_\mu}^{(n)} = \sum_{r=0}^{n-1} v_0^{n-1-r} A_{A_1, A_2, \dots, A_\mu}^{(r)}.$$

This whole procedure, although in principle very simple, may of course be lengthy if m is not small and if no specific assumption for the l.d. is considered; for in the expression of $A_{A_1, A_2, \dots, A_\mu}$ many different α -values appear, - each however with less than m subscripts—which play the role of abbreviations for complicated expressions; in other words the explicit solution for $m = 6$, for instance requires the solutions for $m < 6$, all these solutions being however completely given by our formulae, down to $m = 2$, where $\alpha_{12}^{(n)}$ and $\alpha_{1,2}^{(n)}$ are given by (23).

Under simple assumptions for the l.d. the explicit expressions for the α become simple. Two extreme cases are complete linkage and free assortment. In the first case $p_{12}^{(n)} \dots p_m^{(n)} = p_{12}^{(0)} \dots p_m^{(0)}$ and nothing remains to be done. The case of free assortment where all $v = (\frac{1}{2})^{m-1}$ can be dealt with directly by induction, or we may evaluate the general formulae given above which in this case become quite simple. We have⁸

$$(45) \quad 2^{mn} \alpha_{A_1, A_2, \dots, A_\mu}^{(n)} = 2^n (2^n - 1) \dots (2^n - \mu + 1).$$

That shows that the values of the coefficients $\alpha^{(n)}$ depend only on the number of elements A , which appear as subscripts. Thus we find e.g. for $m = 6$, if we write in each line of (45') one typical value:

$$(45') \quad \begin{aligned} \alpha_{123456}^{(n)} &= 2^n / 2^{6n} = 1 / 2^{5n} \\ \alpha_{1,23456}^{(n)} &= \alpha_{12,3456}^{(n)} = \alpha_{123,456}^{(n)} = (2^n - 1) / 2^{5n} \\ \alpha_{1,2,3456}^{(n)} &= \alpha_{1,23,456}^{(n)} = \alpha_{12,34,56}^{(n)} = (2^n - 1)(2^n - 2) / 2^{5n} \\ \alpha_{1,2,3,456}^{(n)} &= \alpha_{12,34,5,6}^{(n)} = (2^n - 1)(2^n - 2)(2^n - 3) / 2^{5n} \\ \alpha_{1,2,3,4,56}^{(n)} &= (2^n - 1)(2^n - 2) \dots (2^n - 4) / 2^{5n} \\ \alpha_{1,2,3,4,5,6}^{(n)} &= (2^n - 1)(2^n - 2) \dots (2^n - 5) / 2^{5n}. \end{aligned}$$

Thus in the simple case of independent assortment the explicit solution is very simple too. It confirms the fact that $\lim_{n \rightarrow \infty} \alpha_{1,2,3,4,5,6}^{(n)} = 1$ while all other α 's approach

⁸ The values on the right side of (45) are indicated in [1], but the solution for free assortment reported in this article does not seem to coincide with ours.

zero. To prove this, however, without recurring to computations, was the purpose of the preceding section.

9. Crossover distribution and crossover probabilities. The limit theorem of §7 as well as the computations of the preceding section, in short, all investigations and concepts considered so far, are valid for any l.d. We shall now define and use a *crossover distribution*, (c.d.), which is completely equivalent to the l.d. but preferable for the study of certain particular cases. Apparently biologists have not considered the general concept of the c.d. but only the c.p. c_{ij} . This concept is basic and tangible but not sufficient for a complete description of the linkage mechanism when $m \geq 4$, as was seen in the preceding sections.

It is obvious that, from our point of view, a mathematical theory of linkage must be based on the properties of and a set of assumptions on the l.d., or the c.d. The *linear theory* will be considered from this standpoint. This theory is, of course, still compatible with a variety of particular assumptions. In the last section some simple particular cases will be presented and studied with a special view to *interference*.

The probability that an individual transmits the set of "paternal genes" belonging to A and the set of "maternal genes" belonging to $A' = S - A$ is denoted by $l(A)$, where $l(A) = l(A')$; e.g. with $m = 8$: $l(1,0,1,1,1,0,0,1) = l(0,1,0,0,0,1,1,0)$. Considering here the succession of arguments we see that in either set of eight arguments: The first and the second are from different sets, the second and the third are again from different sets, the third and the fourth are from the same set, ... the seventh and eighth are from different sets. Writing 0 for "same" and 1 for "different" and using these numbers to correspond to the $(m - 1)$ consecutive intervals between the m genes, we introduce:

$$l(10111001) + l(01000110) = \pi(1100101).$$

Here $\pi(\eta_1, \eta_2, \dots, \eta_{m-1})$ where $\eta_i = 0$ or 1, is an $(m - 1)$ -variate alternative. The relation between the l.d. and this new distribution may be written in the form

$$(46) \quad 2l(\epsilon_1, \epsilon_2, \dots, \epsilon_m) = \pi(|\epsilon_1 - \epsilon_2|, |\epsilon_2 - \epsilon_3|, \dots, |\epsilon_{m-1} - \epsilon_m|), \quad \epsilon_i = 0 \text{ or } 1.$$

In this definition no fixed "order" of the genes is implied so far. The numbers 1, 2, ... m are used like names.

But it seems to be admitted today by leading biologists that a certain natural order of the genes exists. If this is so the numbers 1, 2, ... m should be used in agreement with this order. Let us note, however, that the situation is in reality slightly different: Only the genes *within each linkage group* (§4) are assumed to be ordered, whereas no order exists among the groups. Let us for the moment disregard this circumstance and assume that all genes under consideration belong to the same linkage group.

Within such a linkage group a one-dimensional or linear order prevails, to be understood in the geometric sense of "location". Some more precise definitions

concerning this linear order will be considered later. For the moment we simply imagine that each of the two sets of genes belonging to an individual is arranged like m consecutive discrete points on a line segment.⁹ The crossover distribution $\pi(\eta_1, \eta_2, \dots, \eta_{m-1})$, introduced in (46) becomes more meaningful under this assumption where, now, the numbering corresponds to this linear order. Then the argument 0 in this distribution can be interpreted as "coherence" and the argument 1 as "interchange" or "crossing over" and the "intervals" as intervals in the geometric sense. Whether this "crossing over", which means transition from the maternal to the paternal set or vice versa, is to be conceived as a "break" (Janssen's chiasmotypie) does not matter for the above definitions. If however, the idea is that between two neighboring genes not more than one break is possible then the "event," which we call crossover, would be at the same time a break; if, biologically, more than one break between i and $(i + 1)$ is not excluded, then the event "crossover within $(i, i + 1)$ " means "odd number of breaks within this interval."

Now, let us consider the relation between the c.d. and the c.p. It has been repeatedly remarked that the c.p. are not equivalent to the l.d., hence they are not equivalent to the c.d. either. There are $\frac{1}{2} \cdot m(m - 1)$ c.p. but $2^{m-1} - 1$ l-values, or π -values. If $m \geq 4$ the second number is greater than the first. Besides, the l-values are absolutely arbitrary probabilities. For the c.p. in section 4 some restrictions were derived. Let us derive another set of restrictions by considering four numbers i, j, k, l which we may denote by 1, 2, 3, 4. (The following computation has nothing to do with linear order. It applies if $m = 4$ to the l.d. $l(\epsilon_1 \epsilon_2 \epsilon_3 \epsilon_4)$ and if $m > 4$ to the respective four-dimensional marginal distributions of the l.d.) Write $v(\epsilon_1 \epsilon_2 \epsilon_3 \epsilon_4) = 2l(\epsilon_1 \epsilon_2 \epsilon_3 \epsilon_4)$ and let us add up the six c.p. corresponding to these four numbers. From $c_{ij} = 2l_{ij}(1, 0) = v_{ij}(1, 0)$ we get

$$\begin{aligned} c_{12} + c_{13} + \dots + c_{34} &= 3v(1000) + 3v(0100) + 3v(0010) + 3v(0001) \\ (47) \quad &+ 4v(1001) + 4v(1010) + 4v(1100) \\ &= 4 - 4v(0000) - v(1000) - v(0100) - v(0010) - v(0001) \leq 4. \end{aligned}$$

Hence as by (14) $c_{12} + c_{23} + c_{13} \leq 2$, it follows that

$$(14') \quad c_{il} + c_{jl} + c_{kl} \leq 2$$

is another necessary condition for the c.p. The limit "2" can be reached, as we see for $v(0000) = v(1000) = v(0100) = v(0010) = v(0001) = 0$; then

$$c_{12} = c_{34} = v(1001) + v(1010)$$

$$c_{23} = c_{14} = v(1100) + v(1010)$$

$$c_{13} = c_{24} = v(1001) + v(1100)$$

⁹ "The genes are represented as lying in a line like beads on a string. The numerical data from crossing over show in fact that this arrangement is the only one that is consistent with the results obtained" [11]. This is but one of many statements in favor of the linear theory

Then $c_{i,i+2}$ is the probability of one interchange between i and $i+2$, i.e. of either an interchange in the first of the two intervals numbered i and $i+1$, and no interchange in the second; or of an interchange in the second but none in the first. Hence $c_{i,i+2} = P_{i,i+1}(1)$, ($i = 1, \dots, n-1$), because $P_{i,i+1}(1)$ is just the probability of exactly one "success" in the two trials numbered i and $i+1$. In the same way we get $c_{i,i+3} = P_{i,i+2}(1) + P_{i,i+2}(3)$, ($i = 1, \dots, n-2$), because an interchange between i and $i+3$ means either exactly one or exactly three interchanges in the three intermediate intervals. Hence we get altogether, with $n = m-1$:

$$\begin{aligned} c_{i,i+1} &= p_i \\ (48) \quad c_{i,i+2} &= P_{i,i+2}(1) \\ c_{1m} &= P_{12\dots n}(1) + P_{12\dots n}(3) + \dots + P_{12\dots n}(\tilde{n}), \text{ where } \tilde{n} = n \text{ if } n \text{ odd,} \\ &= n-1 \text{ if } n \text{ even.} \end{aligned}$$

Let us also express the c_i in terms of the S_i . It is well known (see e.g. [3]) that

$$(49) \quad P_{1\dots n}(x) = \sum_{r=0}^n (-1)^{r+x} S_r \quad (x = 0, 1, \dots, n).$$

Applying these to (48) we easily find the convenient expressions:

$$\begin{aligned} c_{12} &= p_1, \text{ etc.} \\ c_{13} &= (p_1 + p_2) - 2p_{12} \equiv (S_1 - 2S_2)_{12}, \text{ etc.} \\ c_{14} &= (p_1 + p_2 + p_3) - 2(p_{12} + p_{13} + p_{23}) + 4p_{123} \\ (50) \quad &\equiv (S_1 - 2S_2 + 4S_3)_{123}, \text{ etc.} \\ c_{15} &= (S_1 - 2S_2 + 4S_3 - 8S_4)_{1\dots 4}, \text{ etc.} \\ &\dots \\ c_{1,m} &= S_1 - 2S_2 + 4S_3 + \dots + (-2)^m S_{m-1}. \end{aligned}$$

10. The linear theory. Consider a linkage group of size m and assume for the moment that $c_{i,j} \neq c_{i,k}$ for all i, j , and k . It seems that the main mathematical content of the linear theory can be summarized as follows: *It is possible to establish in a unique way an order or a succession of the genes, such that for the*

$$\binom{m}{2} = \frac{m(m-1)}{2} \dots \text{c.p. the } (m-1)(m-2) \text{ inequalities}$$

$$(51) \quad \begin{aligned} c_{i,j} &< c_{i,j+1} & (i = 1, 2, \dots, m-2) \\ c_{i,j} &< c_{i-1,j} & (i = 2, 3, \dots, m-1) \end{aligned} \quad (i < j)$$

hold. In this succession j will be between i and k if $c_{i,k}$ is greater than the two other c.p. $c_{i,j}$ and $c_{j,k}$. The two arrangements $1, 2, \dots, m$ and $m, m-1, \dots,$

$\dots 1$ are considered as corresponding to the same order. Furthermore, this order is a straight-line-succession for which an additive distance relation holds (cf. also [4a] and [13]). Instead of the restriction $c_{i,j} \neq c_{i,k}$ it is sufficient to assume the weaker restriction, that in any triple $c_{i,j}$, $c_{i,k}$, $c_{j,k}$, one is greater than the two others. Without such a restriction uniqueness of the order no longer holds. E.g. in case of independent assortment where all $c_{i,j}$ equal $\frac{1}{2}$ any of the m possible numberings of the genes is equally admissible from the point of view of the linear theory. In the case of complete linkage where all c.p. are zero it will be logical to consider all m genes as located in the same point. Obviously there are all kinds of intermediate cases. We shall come back to this point at the end of this section.

Now consider again the case of "different" c.p. (in the above defined sense). Let us prove that there can be not more than one succession for which (51) holds. In fact it follows from (51) that also:

$$(51') \quad c_{ij} < c_{ik} \quad (i = 1, 2, \dots, m-2) \quad \text{for all } k > j \quad i < j$$

and

$$c_{ij} < c_{kj} \quad (i = 2, 3, \dots, m-1) \quad \text{for all } k < i.$$

These are all together $M = 2 \cdot 1 + 3 \cdot 2 + \dots + (m-1)(m-2) = 2 \cdot \binom{m}{3}$

inequalities. On the other hand there are all together $\binom{m}{3} = M/2$ "between"-relations for m numbers, each of them being defined by two inequalities as $c_{i,j} < c_{i,k}$ and $c_{j,k} < c_{i,k}$ (if j is between i and k); hence on the whole M such inequalities. But these are the same as (51'), as we see by changing i, j, k into j, k, i in the second equation (51'). Thus it is not possible to find two different successions which both satisfy (51).

As to the *metric* of the problem, Morgan proposed originally that the value of the c.p. $c_{i,j}$ should be used as the distance between i and j . It has, however been objected repeatedly that this distance would not be additive; this is obvious since the triangular relation (13) holds for three subscripts (see also (50)).¹¹ The equality $c_{i,j} + c_{j,k} = c_{i,k}$ holds only in the exceptional cases where multiple crossingover is excluded. It seems, however that an adequate definition of distance is available if we try to formulate in terms of probability theory what the biologist had in mind. Let $j \geq i$. The distance $d_{i,j+1}$ between i and $j+1$ may be defined as the mathematical expectation of the number of crossingovers in $(i, j+1)$, i.e. in the $j+1-i$ intervals between i and $j+1$. Hence if

¹¹ For a geometric equivalent of m points with $m(m-1)/2$ arbitrary distances we would have to turn to an $(m-1)$ -dimensional space. In fact it is well known that there are between k points in the plane only $S_2 = 2k-3$ arbitrary distances, in space only $S_3 = 3k-6$, in r -space $S_r = rk - r(r+1)/2$. Hence for $r = m-1$ and $k = m$: $S_{m-1} = m(m-1)/2$.

$P_{1,\dots,j}(x)$ denotes, as before, the probability of exactly x crossovers in the $(j+1-i)$ intermediate intervals the formula holds

$$(52) \quad d_{i,j+1} = \sum_{x=0}^{j+1-i} x P_{1,\dots,j}(x).$$

Of course, an appropriate unit may be used such that in practical use the distance becomes *proportional* to the d_i , introduced above.

The mean value to the right in (52) is well known for any distribution $\pi(x_1, \dots, x_n)$ whether an "independent" or a general distribution; (i.e. in our case: with or without "interference"). Denoting in the usual way by $\pi_i(x_i)$ the marginal distributions of first order of $\pi(x_1, \dots, x_n)$ and putting $\pi_i(1) = p_i$ = the probability of success in the i -th trial, we get:

$$(53) \quad d_{i,j+1} = p_i + p_{i+1} + \dots + p_j,$$

and in the same way with $k > j$

$$d_{j+1,k+1} = p_{j+1} + p_{j+2} + \dots + p_k$$

$$d_{i,k+1} = p_i + p_{i+1} + \dots + p_k$$

hence $d_{i,j+1} + d_{j+1,k+1} = d_{i,k+1}$, or in general:

$$(54) \quad d_{ij} + d_{jk} = d_{ik} \quad (i < j < k).$$

It may be mentioned that the additive property of the mathematical expectation which was used here is very well known (particularly for independent events) but not always correctly proved. The proof is contained in the transformation expressed in the following equalities:

$$\begin{aligned} d_{i,j+1} &= \sum_{x=0}^{j+1-i} x P_{1,\dots,j}(x) \\ (55) \quad &= \sum_{x_i} \sum_{x_{i+1}} \dots \sum_{x_j} (x_i + x_{i+1} + \dots + x_j) \pi_{1,i+1,\dots,j}(x_1, \dots, x_j) \\ &= \sum_{x_1} \sum_{x_2} \dots \sum_{x_n} (x_i + x_{i+1} + \dots + x_j) \pi(x_1, x_2, \dots, x_n). \end{aligned}$$

(For general distributions Stieltjes integrals replace the sums.) In (55) $\pi(x_1, \dots, x_n)$ is the given n -variate distribution, $P_{1,\dots,j}$ the probability of exactly x successes in the successive trials numbered i, \dots, j and $\pi_{1,i+1,\dots,j}(x_1, \dots, x_j)$ is the respective marginal distribution of $\pi(x_1, \dots, x_n)$. The first equality in (55) is not obvious, while the second is rather trivial. From the second or third form of $d_{i,j+1}$ in (55), follows (53). The last expression in (55) shows that the expectation of any such sum as $(x_i + x_{i+1} + \dots + x_j)$ can be computed with respect to one and the same distribution $\pi(x_1, \dots, x_n)$. Therefore the distance $d_{i,j+1}$ may also be defined as the expectation of $(x_i + x_{i+1} + \dots + x_j)$ with respect to the c.d.

Because of the first equation (48) we get from (53)

$$(53') \quad d_{i,j} = c_{i,i+1} + c_{i+1,i+2} + \cdots + c_{j-1,j}.$$

Hence the distance $d_{i,j}$ is equal to the sum of the $j - i$ intermediate c.p. No difficulty arises for us from the obvious fact that always

$$(53'') \quad c_{i,j} \leq d_{i,j}; \text{ and in general } c_{i,j} < d_{i,j},$$

because the distance $d_{i,j}$ is defined by (52), or (55) and not as $c_{i,j}$.

On the right side in (53') stands the sum of certain c.p. We have repeatedly remarked that there may be hitherto unknown restrictions for a consistent system of c.p. Hence it is important to notice that *there are no restrictions for the particular $(m - 1)$ c.p. $c_{12}, c_{23}, \dots, c_{m-1,m}$. They can be quite arbitrarily chosen because of $c_{i,i+1} = p_i$. Hence any geometric representation of m genes arranged on a straight line in arbitrary distances $d_{i,i+1}$ ($i = 1, 2, \dots, m - 1$) is surely consistent. E.g. m consecutive genes may be arranged with equal distances $d_{12} = d_{23} = \dots = d_{m-1,m}$. Or some distances may be zero; then the respective genes are localized in the same point, etc.*

Finally, let us briefly consider the case of *several linkage groups*. According to §4 the l.d. then resolves into a product of several distributions; e.g. with $m = 12$:

$$(56) \quad \begin{aligned} l(\epsilon_1 \epsilon_2, \dots, \epsilon_{12}) &= f_1(\epsilon_1 \epsilon_2 \epsilon_3 \epsilon_4) f_2(\epsilon_5 \epsilon_6 \epsilon_7) f_3(\epsilon_8 \epsilon_9 \epsilon_{10}) f_4(\epsilon_{11} \epsilon_{12}) \\ &= \left(\frac{1}{2}\right)^4 \pi_1(|\epsilon_1 - \epsilon_2|, |\epsilon_2 - \epsilon_3|, |\epsilon_3 - \epsilon_4|) \\ &\quad \pi_2(|\epsilon_5 - \epsilon_6|, |\epsilon_6 - \epsilon_7|) \cdots \pi_4(|\epsilon_{11} - \epsilon_{12}|). \end{aligned}$$

Then, as postulated by Morgan, the linear order holds within each of the k groups, whereas all c.p. among the groups are equal to $\frac{1}{2}$.

Let us conclude this section by transforming the basic conditions (51) of the linear theory by means of (48). This will be needed in the following section. Consider e.g. the condition $c_{13} < c_{14}$, i.e.

$$(57') \quad P_{12}(1) < P_{123}(1) + P_{123}(3)$$

or $c_{24} < c_{14}$ yields $P_{23}(1) < P_{123}(1) + P_{123}(3)$. Or in the same way:

$$(57'') \quad \begin{aligned} P_{123}(1) + P_{123}(3) &< P_{1234}(1) + P_{1234}(3) \\ &< P_{1, \dots, 5}(1) + P_{1, \dots, 5}(3) + P_{1, \dots, 5}(5), \text{ etc.} \end{aligned}$$

Thus we may express the content of (51) as follows: *The probability that the "event" happens an odd number of times in a set T_i of i consecutive trials is less than the probability that the event happens an odd number of times in the set T'_{i+1} or in the set T'_{i+1} each consisting of $i + 1$ consecutive trials where T_{i+1} and T'_{i+1} denotes respectively the sum of T_i and either the immediately following or the immediately preceding trial* In this form we see again that the linear theory is an assumption, suggested by observations, and by no means logically necessary.

11. Some models of c.d.'s based on the linear theory. The simplest and very important example which has been suggested repeatedly is that of independent crossovers:

(i) *Independence.* The crossovers do not influence each other, i.e.

$$(58) \quad p_{i,j} = p_i p_j, \quad p_{i,j,k} = p_i p_j p_k, \dots$$

That this distribution is consistent is well known; hence only the specific inequalities (48) or (57) have to be considered. Here the expressions $P_{12} \dots (x)$ used in (57) become very simple, e.g. with $p_i + q_i = 1$:

$$P_{1,2}(1) = p_1 q_2 q_3 q_4 + q_1 p_2 q_3 q_4 + q_1 q_2 p_3 q_4 + q_1 q_2 q_3 p_4.$$

Then a simple computation shows:

$$(59) \quad \begin{aligned} c_{i,j+1} - c_{i,j} &= (q_i - p_i) \cdots (q_{j-1} - p_{j-1}) p_j \\ c_{i-1,j} - c_{i,j} &= p_{i-1} (q_i - p_i) \cdots (q_{j-1} - p_{j-1}). \end{aligned}$$

These differences will be positive if all $q_i - p_i > 0$ or all $p_i < \frac{1}{2}$. Hence: A consistent c.d. which fulfils the conditions (51) of the linear theory is the distribution of "independent crossovers" with basic probabilities $p_i = c_{i,i+1}$ ($i = 1, 2, \dots, m-1$), with the one restriction

$$(60) \quad c_{i,i+1} = p_i \leq \frac{1}{2}.$$

The distribution is completely determined by (58). If all $p_i = p = \frac{1}{2}$, we have the particular case of free assortment.

Although this independence is more general than Mendel's original assumption, Morgan, Haldane and others reported observations, not in accordance with this hypothesis. One crossingover seems to prevent others in a certain "neighborhood". This phenomenon was named *interference*. It suggests that we have to consider the c.d. as a distribution of dependent rather than of independent events. This will be done in the following pages. First consider the limit-case of:

(ii) *Complete interference or disjoint events.* In this case we have

$$(61) \quad p_{i,j} = p_{i,j,k} = \dots = p_{i2,\dots,m-1} = 0.$$

Thence it follows that we have simply

$$(62) \quad \begin{aligned} c_{i,i+1} &= p_i \\ c_{i,i+2} &= p_i + p_{i+1} \\ c_{i,i+3} &= p_i + p_{i+1} + p_{i+2}, \text{ etc.} \end{aligned}$$

In this particular case the c.p. are additive $c_{i,j} = d_{i,j}$. It is obvious from (62) that in this case the conditions (51) of the linear theory are fulfilled. On the other hand it follows from (49), (for $x = 0$) and (61) that the system is consistent if and only if

$$(63) \quad S_1 = p_1 + p_2 + \dots + p_n \leq 1 \quad (n = m - 1).$$

This is in accordance with the fact that nearly or exactly additive c.p. have been observed always in connection with very small p_i -values.

The most striking observation leading to the concept of interference was that $p_{ij} \leq p_i p_j$, i.e. that double crossovers appeared less frequently than one would have assumed for independent crossovers, but that nevertheless they did appear sometimes. A particularly simple model of dependence or interference which starts with this fact, preserving however, the main structure of independence, is the following:

(iii) *One-parametric model of partial interference.* Assume as before ($m - 1$) basic probabilities p_i and put

$$(64) \quad p_{ij} = \epsilon p_i p_j, \quad p_{ijk} = \epsilon p_i p_j p_k, \dots, \text{etc.} \quad (0 \leq \epsilon \leq 1).$$

There is independence if $\epsilon = 1$, complete interference for $\epsilon = 0$ and partial interference for intermediate values of ϵ . Let us first investigate conditions for the consistency of this distribution. Necessary and sufficient conditions for a consistent distribution of arbitrarily linked events are well known (see e.g. [3] (b) p. 239). Write $m - 1 = n$. A system of $p_i, p_{ij}, \dots, p_{1\dots n}$ is consistent if it is possible to compute from these $(2^n - 1)$ values, 2^n non-negative values $\pi(\eta_1, \eta_2, \dots, \eta_n)$ ($\eta_i = 0$ or 1) which have the sum one and are given by the formulae:

$$\begin{aligned} \pi(11\dots 1) &= p_{12\dots n} \\ \pi(11\dots 10) &= p_{12\dots (n-1)} - p_{12\dots n} \\ &\dots\dots\dots \\ (65) \quad \pi(110\dots 0) &= p_{12} - \sum_{n_1} p_{12n_1} + \sum_{n_1} \sum_{n_2} p_{12n_1 n_2} - \dots \pm p_{12\dots n} \\ &\dots\dots\dots \\ \pi(00\dots 0) &= 1 - \sum_{n_1} p_{n_1} + \sum_{n_1} \sum_{n_2} p_{n_1 n_2} - \dots \pm p_{12\dots n}. \end{aligned}$$

Because of the symmetry of (64) it will be sufficient to check (65) by means of the relations (49) which can be obtained from (65) by collecting groups of equations such that the corresponding $\pi(\eta_1, \dots, \eta_n)$ show all the same number of 1's as arguments. Write $P_{1\dots n}(x) = P_x$ for the probability of x successes in the n trials where independent events with basic probabilities p_i are considered, and P'_x for the analogous probability corresponding to the distribution (64) and introduce in the same way S_i and S'_i where as before, $S_i = \sum p_{ij}$, $S_2 = \sum p_{ijk}$, etc. We then find:

$$\begin{aligned} (66) \quad P'_0 &= 1 - S_1 + \epsilon S_2 - \epsilon S_3 + \dots \\ &= P_0 + (1 - \epsilon)(-S_2 + S_3 - S_4 + \dots) = P_0\epsilon + (1 - S_1)(1 - \epsilon). \end{aligned}$$

It follows that the expression $P_0\epsilon + (1 - S_1)(1 - \epsilon)$ must be ≥ 0 . For $\epsilon = 0$ this condition reduces to (63), whereas for $\epsilon = 1$, there is no restriction at all. On the other hand this is the only restriction of this kind, because we find

$$\begin{aligned} P'_1 &= S'_1 - 2S'_2 + 3S'_3 - 4S'_4 + \dots = S_1 - 2\epsilon S_2 + 3\epsilon S_3 - 4\epsilon S_4 + \dots \\ &= P_1 + (1 - \epsilon)(2S_2 - 3S_3 + 4S_4 \dots) = P_1\epsilon + S_1(1 - \epsilon). \end{aligned}$$

This last expression is always ≥ 0 because of $P_1 \geq 0$, $S_1 \geq 0$, $\epsilon \leq 1$. Furthermore we find for $i \geq 2$ that $P_i^* = P_i \epsilon$, hence always non negative. Therefore our system is consistent under the one condition (66).

The additional restrictions corresponding to the linear theory have still to be considered. A simple computation yields the result

$$(67) \quad \begin{aligned} c_{i,j+1} - c_{ij} &= (1 - 2\epsilon p_1)(1 - 2\epsilon p_{i+1}) \cdots (1 - 2\epsilon p_{j-1})p_j \\ c_{i,j+1} - c_{i+1,j+1} &= p_i(1 - 2\epsilon p_{i+1}) \cdots (1 - 2\epsilon p_j). \end{aligned}$$

These differences are ≥ 0 if $p_i \leq \frac{1}{2\epsilon}$ which is, for $\epsilon < 1$, less strong than (60).

Hence we sum up: A consistent model of partial interference with one parameter ϵ to fit the observations can be obtained on the basis of $n = m - 1$ probabilities p_1, p_2, \dots, p_n by means of (64), if the condition

$$(68) \quad P_0 \epsilon + (1 - S_1)(1 - \epsilon) \geq 0 \quad \text{or} \quad S_1 \leq 1 + \frac{\epsilon}{1 - \epsilon} P_0$$

holds and the additional restriction required by the "linear theory"

$$(69) \quad p_i \leq \frac{1}{2\epsilon}$$

is satisfied. For $\epsilon = 1$ this reduces to "independent events" or "no interference" with no restriction (68), and (69) reducing to (60). For $\epsilon = 0$ our model yields "complete interference" or "disjoint events" with restriction (68) becoming (63) and no restriction (69). If we say that this model contains one parameter only, the idea is that the p_i are to be identified with the basic c.p. $c_{i,i+1}$. It might, however, seem adequate to consider ϵ and p_1, \dots, p_{m-1} as m available parameters which may be determined from the observations by some appropriate method.

(iv) An $(m - 1)$ -parametric model of partial interference. Numerical data show (see particularly [4]) that interference is particularly marked i.e. $p_i < p_i p_j$, if the corresponding p_i, p_j are very small, whereas for greater values of the p_i we have more nearly the pattern of independence. This is rather a striking fact, and seems to be well confirmed by observation. In these final pages a model will be studied which takes into account the circumstance that the amount of interference seems to depend on the magnitudes of the p_i . It contains $(m - 1)$ parameters, is therefore rather flexible, but nevertheless very simple.

Assume $m - 1 = n$ numbers ϵ_i where $0 \leq \epsilon_i \leq 1$ and form by means of n probabilities p_i :

$$(70) \quad \epsilon_i p_i = \bar{p}_i \quad (0 \leq \epsilon_i \leq 1) \quad (i = 1, 2, \dots, n).$$

We may choose ϵ_i small if the corresponding p_i is small and larger if it is large; if the p_i 's are all of the same order of magnitude the ϵ_i 's need not differ much either. Then we simply define:

$$(71) \quad p_i = \bar{p}_i \bar{p}_j, \quad p_{ijk} = \bar{p}_i \bar{p}_j \bar{p}_k, \dots, \quad p_{12\dots n} = \bar{p}_1 \bar{p}_2 \dots \bar{p}_n.$$

Let us investigate the consistency of this model. In analogy to (66) we form with $S_1 = \sum p_i$, $\bar{S}_1 = \sum \bar{p}_i$, $\bar{S}_2 = \sum \bar{p}_i \bar{p}_j$, etc.:

$$(72) \quad \begin{aligned} P'_0 &= 1 - S_1 + \bar{S}_2 - \bar{S}_3 + \cdots \\ &= (1 - \bar{S}_1 + \bar{S}_2 - \bar{S}_3 + \cdots) - \sum_{i=1}^n (1 - \epsilon_i) p_i = \bar{P}_0 - \sum_{i=1}^n (1 - \epsilon_i) p_i \end{aligned}$$

where P'_0 and \bar{P}_0 are the probabilities for zero successes for the model under consideration and for independent events with basic probabilities \bar{p}_i , respectively;

hence $\bar{P}_0 = \prod_{i=1}^n (1 - \epsilon_i p_i)$ and we get the condition:

$$(73) \quad \prod_{i=1}^n (1 - \epsilon_i p_i) \geq \prod_{i=1}^n (1 - \epsilon_i) p_i \quad \text{or:} \quad \sum_{i=1}^n p_i \leq \sum_{i=1}^n \bar{p}_i + \prod_{i=1}^n (1 - \bar{p}_i).$$

If all $\epsilon_i = 1$ there is no restriction (73), while for $\epsilon_i = 0$ we find again (63). The consideration of P'_1, P'_2, \dots yields no new condition, because we get, denoting by \bar{P}_i the probability of i successes for the independent events with basic probabilities \bar{p}_i :

$$\begin{aligned} P'_1 &= S_1 - 2\bar{S}_2 + 3\bar{S}_3 - \cdots \pm n\bar{S}_n = \bar{S}_1 - 2\bar{S}_2 + \cdots \pm n\bar{S}_n + \sum_{i=1}^n p_i(1 - \epsilon_i) \\ &= \bar{P}_1 + \sum_{i=1}^n p_i(1 - \epsilon_i) \geq 0 \quad \text{and:} \end{aligned}$$

$$P'_i = \bar{P}_i \geq 0 \quad (i \geq 2).$$

As for the restrictions imposed by the linear theory we find:

$$(74) \quad \begin{aligned} c_{i,j+1} - c_{i,j} &= (1 - 2\bar{p}_i)(1 - 2\bar{p}_{i+1}) \cdots (1 - 2\bar{p}_{j-1})\bar{p}_j + p_j(1 - \epsilon_j) \\ c_{i,j+1} - c_{i+1,j+1} &= \bar{p}_i(1 - 2\bar{p}_{i+1}) \cdots (1 - 2\bar{p}_j) + p_i(1 - \epsilon_i). \end{aligned}$$

Thus the conditions of the linear theory are satisfied if

$$(75) \quad \bar{p}_i \leq \frac{1}{2} \quad \text{or} \quad p_i \leq \frac{1}{2\epsilon_i}.$$

Hence summarizing: *On the basis of $m - 1$ probabilities p_i a consistent model of partial interference is obtained by means of (70) and (71) if the condition of consistency (73) and the conditions (75) are satisfied.*

It may be that the four simple models described in this section will seem too crude for the description of the complex mechanism of linkage. They could, of course, be combined and modified in various ways in order to serve at least as an approximation to the theoretical picture of reality we wish to construct. But, while these particular attempts may be inadequate, it seems to the author that the underlying principle is not wrong: that a mathematical theory of linkage must finally consist in statements on the l.d. (or the equivalent c.d.). The consideration of the c.p. is not sufficient for this purpose. The mathematical instrument for a theory of linkage seems to be the probability theory of the linkage distribution.

REFERENCES

- [1] F. BERNSTEIN, *Variations- und Erblichkeitsstatistik*. Handbuch der Vererbungswissenschaft, Bd 1, pp 1-96.
- [2] KAI LAI CHUNG, "On fundamental systems of probabilities of a finite number of events" *Ann. of Math. Stat.*, Vol. 14 (1943), pp. 234-37.
- [3] H. GEIRINGER, (a) On the probability theory of arbitrarily linked events *Ann. of Math. Stat.*, Vol. 9 (1938), pp. 260-271.
 (b) "A note on the probability of arbitrary events," *Ann. of Math. Stat.*, Vol. 13 (1942), pp. 238-245.
- [4] J. B. S. HALDANE, (a) "The combination of linkage values and the calculation of distances between the loci of linked factors," *Jour. of Genetics*, Vol. 8 (1919), pp. 299-308.
 (b) "Theoretical genetics of autopolyploids," *Jour. of Genetics*, Vol. 22 (1930), pp. 359-372.
- [5] G. H. HARDY, "Mendelian proportions in a mixed population," *Science*, Vol. 28, (1908), p. 49-50.
- [6] J. S. HUXLEY (editor), *The New Systematics*, Oxford 1940. (See articles by S. Wright, p 161-183 and H. J. Muller, p. 158-268).
- [7] H. S. JENNINGS, (a) "The numerical results of diverse systems of breeding with respect to two pairs of characters, etc.," *Genetics*, Vol. 12 (1917), pp. 97-154.
 (b) "The numerical relations in the crossing over of the genes with a critical examination of the theory that the genes are arranged in a linear series," *Genetics*, Vol. 8 (1923), p. 303.
- [8] K. v. KÖRÖSY, *Versuch einer Theorie der Genkoppelung*. (Bibliotheca Genetica), Leipzig 1920.
- [9] K. MATHER, *The Measurement of Linkage in Heredity*, London 1938.
- [10] G. MENDEL, "Versuche über Pflanzenhybriden." *Verh. des Naturforsch. Vereines in Brünn*, IV. Bd., *Abhandlungen Brünn*, 1866 pp. 3-47.
- [11] T. H. MORGAN, *The Theory of the Gene*. New Haven, 1928.
- [12] K. PEARSON, "On a generalized theory of alternative inheritance with special reference to Mendel's laws," *Phil. Trans. Royal Soc. (A)*, Vol. 203 (1904), pp. 33-86.
- [13] H. RADEMACHER, "Mathematische Theorie der Genkoppelung unter Berücksichtigung der Interferenz", 105. *Jahresber. (1932)*, Schles. Ges. f. vaterländische Kultur pp. 1-8.
- [14] R. B. ROBBINS, (a) "Applications of mathematics to breeding problems, II." *Genetics*, Vol. 3 (1918), pp. 73-92.
 (b) "Some applications of mathematics to breeding problems III." *Genetics*, Vol. 3 (1918), pp. 375-389.
- [15] H. TIETZE, "Über das Schicksal gemischter Populationen nach den Mendelschen Vererbungsgesetzen," *Zs. Angew. Math. u. Mech.*, Bd. 3, (1923), pp. 362-393.
- [16] W. WEINBERG, "Über Vererbungsgesetze beim Menschen." *Zs. f. induktive Abstammungs- und Vererbungslehre*, Vol. 1 (1909) p. 277-330.
- [17] S. WRIGHT, "Statistical genetics and evolution," *Bull. Amer. Math. Soc.*, Vol. 48 (1942), pp. 223-246.

THE COVARIANCE MATRIX OF RUNS UP AND DOWN

BY H. LEVENE AND J. WOLFOWITZ

Columbia University

1. Introduction. Let a_1, \dots, a_n be n unequal numbers and let the sequence $S = (h_1, h_2, \dots, h_n)$ be any permutation of a_1, \dots, a_n . S is to be considered a chance variable, and each of the $n!$ permutations of a_1, \dots, a_n is assigned the same probability. Consider the sequence R whose i^{th} element is the sign (+ or -) of $h_{i+1} - h_i$, ($i = 1, 2, \dots, n-1$). A sequence of p consecutive + signs not immediately preceded or followed by a + sign is called a run up of length p ; a sequence of p consecutive - signs not immediately preceded or followed by a - sign is called a run down of length p . The term "run" will denote both runs up and runs down. The usage of the term "length" varies; most quality control literature attributes the length $p+1$ to the runs which we say are of length p .

As an example of our usage, the sequence

$$S = 2 \ 8 \ 13 \ 1 \ 3 \ 4 \ 7$$

gives the sequence

$$R = + \ + \ - \ + \ + \ +,$$

which has a run up of length 2, followed by a run down of length 1, followed by a run up of length 3.

Runs up and down are widely used in quality control and have been applied to economic time series. The purpose of this paper is to obtain their variances and covariances and to correct some erroneous notions prevalent in the literature about their application.

2. Notation. If the sign (+ or -) of $(h_{i+1} - h_i)$ is the initial sign of a run defined as above, we call h_i the initial turning point (i. t. p.) of the run. Then h_i is always an i. t. p., and we adopt the convention that h_n is never an i. t. p. We define new stochastic variables as follows:

$$(2.1) \quad x_i = \begin{cases} 1 & \text{if } h_i \text{ is an i. t. p.,} \\ 0 & \text{otherwise,} \end{cases}$$

$$(2.2) \quad x_{pi} = \begin{cases} 1 & \text{if } h_i \text{ is the i. t. p. of a run of length } p, \\ 0 & \text{otherwise,} \end{cases}$$

$$(2.3) \quad w_{pi} = \begin{cases} 1 & \text{if } h_i \text{ is the i. t. p. of a run of length } p \text{ or more,} \\ 0 & \text{otherwise,} \end{cases}$$

for $i = 1, 2, \dots, n$. Also

$$(2.4) \quad r = \text{the number of runs in } R,$$

$$(2.5) \quad r_p = \text{the number of runs of length } p \text{ in } R,$$

$$(2.6) \quad r'_p = \text{the number of runs of length } p \text{ or more in } R.$$

Evidently $r = \sum_{i=1}^n x_i$, $r_p = \sum_{i=1}^n x_{pi}$, and $r'_p = \sum_{i=1}^n w_{pi}$.

If X and Y are stochastic variables, let $E(X)$ denote the mean of X , $\sigma(XY)$ denote the covariance of X and Y , and $\sigma^2(X)$ denote the variance of X , if they exist. By the distribution function of X we shall mean a function $f(x)$ such that $P\{X \leq x\} = f(x)$, where the symbol $P\{\cdot\}$ denotes the probability of the relation in the brackets.

3. Preliminary formulas. Let Y be a stochastic variable with any continuous distribution function $f(y)$. Let Y_1, Y_2, \dots, Y_n be a sequence of n independent observations on Y . Since $P\{y_1, \dots, y_n\} = 0$, $0 \neq y_j$, $j = 1, 2, \dots, n$, the distribution of any y_j and Y in Y is evidently the same as that in S . Now choose $f(y) = 1$:

$$\begin{aligned} f(y) &= 0, & y < 0, \\ f(y) &= y, & 0 \leq y < 1, \\ f(y) &= 1, & y \geq 1. \end{aligned}$$

Then

$$P\{y_1 \leq y_2 \leq \dots \leq y_n\} = \int_0^1 \left[\int_{y_{n-1}}^1 \left(\int_{y_1}^{y_2} dy_{n-1} \right) du \right] dy_{n-1} = \frac{1}{3}.$$

By symmetry

$$P\{y_1 \leq P\{y_1 \leq y_2 \leq y_3 \leq \dots \leq y_{n-1} \leq y_n\} = \frac{1}{3},$$

$$n = 2, 3, \dots, n-1).$$

Also $P\{y_1 \leq 1\}$ and $P\{y_n \leq 0\} = 0$.

It will be necessary hereafter to evaluate expressions of the type

$$(3.1) \quad V = \int_0^{y_{p+1}} \int_0^{y_1} \frac{y_1^k}{k!} dy_1 \dots dy_p = \frac{(y_{p+1})^{k+p}}{(k+p)!},$$

and

$$(3.2) \quad V = \int_{y_{p+1}}^1 \int_0^{y_1} \frac{(y_1)^k}{k!} dy_1 \dots dy_p.$$

From the fact that

$$\int_{y_{p+1}}^1 \int_0^{y_1} (y_1) \dots dy_p = \int_0^{y_{p+1}} \dots \int_0^1 dx_1 \dots dx_p = \frac{(y_{p+1})^p}{p!},$$

where $v_i = (1 - y_{p+1})$, $i = 1, \dots, p+1$, it can easily be shown that

$$\begin{aligned} (3.3) \quad V &= \sum_{r=0}^p (-1)^{r+1} \frac{(y_{p+1})^{k+p}}{(k+r)(p-r)!} + (-1)^p \frac{(y_{p+1})^{k+p}}{(k+p)!} \\ &= \sum_{r=0}^p (-1)^{r+1} \frac{(v_{p+1})^{k+p-r}}{(p+k-r)! r!}. \end{aligned}$$

We shall also need $\int_0^1 V dy_{p+1}$ and $\int_0^1 \int_0^{y_{p+1}} V dy_{p+1} dy_{p+2}$. Now

$$(3.4) \quad \int_0^1 V dy_{p+1} = \sum_{r=0}^p (-1)^{r+1} \frac{1}{p+k-r+1! r!}.$$

Making use of the relation,

$$\sum_{r=0}^l (-1)^r \frac{1}{(n-r)!r!} = (-1)^l \frac{1}{n(n-l-1)!l!}, \quad (l < n),$$

we have

$$(3.5) \quad \int_0^1 V dy_{p+1} = \frac{1}{(p+k+1)p!k!}.$$

Similarly

$$(3.6) \quad \int_0^1 \int_0^{y_{p+2}} V dy_{p+1} dy_{p+2} = \frac{1}{(p+k+1)p!k!} - \frac{1}{(p+k+2)(p+1)!k!}.$$

4. Covariances of runs up and down. We first compute $E(r_p)$ and $E(r'_p)$. We define the symbol

$$P\{-, +^p, -\} = P\{y_{i-1} > y_i < y_{i+1} < \dots < y_{i+p} > y_{i+p+1}\}.$$

The value of the right member is independent of i whenever it is defined (i.e., $i-1 \geq 1, i+p+1 \leq n$). Now

$$\begin{aligned} E(x_{pi}) &= P\{-, +^p, -\} + P\{+, -^p, +\} = 2P\{-, +^p, -\} \\ &= 2 \int_0^1 \int_{y_{i+p+1}}^1 \int_0^{y_{i+p}} \dots \int_0^{y_{i+1}} \int_{y_i}^1 dy_{i-1} \dots dy_{i+p+1} = 2 \frac{p^3 + 3p + 1}{(p+3)!}, \\ &\quad (i = 2, 3, \dots, n-p-1). \end{aligned}$$

$$E(x_{p1}) = 2P\{+^p, -\} \quad \text{and} \quad E(x_{p,n-p}) = 2P\{-, +^p\}.$$

By symmetry $E(x_{p1}) = E(x_{p,n-p})$, the common value being $2 \frac{p+1}{(p+2)!}$. Also $E(x_{pi}) = 0$, ($i > n-p$). Thus

$$\begin{aligned} E(r_p) &= E\left(\sum_{i=1}^n x_{pi}\right) = 2E(x_{p1}) + (n-p-2)E(x_{pi}) \\ (4.1) \quad &= 2n \frac{p^3 + 3p + 1}{(p+3)!} - 2 \frac{p^3 + 3p^2 - p - 4}{(p+3)!}, \quad (p \leq n-2). \end{aligned}$$

Besson [1], Kermack and McKendrick [5], and Wallis and Moore [6] gave the exact value, although Besson proved it only for special cases. R. A. Fisher [3] gave $\lim_{n \rightarrow \infty} \frac{E(r_p)}{E(r)}$.

It is clear that $E(w_{pi}) = E(x_{p,n-p})$, ($i = 2, \dots, n-p$), while $E(w_{p1}) = 2P\{+^p\} = 2/(p-1)!$. We then have

$$(4.2) \quad E(r'_p) = 2n \frac{p+1}{(p+2)!} - 2 \frac{p^2 + p - 1}{(p+2)!}, \quad (p \leq n-1).$$

Setting $p = 1$ we have

$$(4.3) \quad E(r'_1) = E(r) = \frac{1}{2}(2n - 1).$$

Formula (4.3) was given by Bienaymé [2].

We now obtain $\sigma(r_p, r_q)$. Let $(x_{p_i} = E(x_{p_i})) = x_{p_i}$. Then

$$(4.4) \quad \begin{aligned} \sigma(r_p, r_q) &= E\left\{\left[\sum_{i=1}^n x_{p_i}\right]\left[\sum_{i=1}^n x_{q_i}\right]\right\} \\ &= \sum_i E(x_{p_i} x_{q_i}) + \sum_{i < j} \sum E(x_{q_i} x_{p_j}) + \sum_{i < j} \sum E(x_{p_i} x_{q_j}). \end{aligned}$$

For $j \geq i + q + 3$, x_{q_i} and x_{p_j} are independent and hence $E(x_{q_i} x_{p_j}) = 0$. Omitting zero terms from (4.4) we have

$$(4.5) \quad \begin{aligned} \sigma(r_p, r_q) &= \left\{ \sum_i E(x_{p_i} x_{q_i}) + \sum_{i < j < i+q+3} E(x_{q_i} x_{p_j}) + \sum_{i < j < i+p+3} E(x_{p_i} x_{q_j}) \right. \\ &\quad - \left\{ \sum_i E(x_{p_i}) E(x_{q_i}) + \sum_{i < j < i+q+3} E(x_{q_i}) E(x_{p_j}) \right. \\ &\quad \left. \left. + \sum_{i < j < i+p+3} E(x_{p_i}) E(x_{q_j}) \right\} \right\}. \end{aligned}$$

Since $x_{p_i} x_{p_i} = \delta_{pq}(x_{p_i})^2 = \delta_{pq} x_{p_i}$, we have for the first term of the right member of (4.5)

$$(4.6) \quad \sum_{i=1}^n E(x_{p_i} x_{q_i}) = \delta_{pq} E(r_p),$$

where the Kronecker delta $\delta_{pq} = \begin{cases} 1, & \text{if } p = q, \\ 0, & \text{otherwise.} \end{cases}$

Since $x_{q_i} x_{p_j} = 0$ for $i < j < i + q$, the second term in the right member of (4.5) is

$$(4.7) \quad \begin{aligned} &\sum_{i=1}^{n-p-q} E(x_{q_i} x_{p, i+q}) + \sum_{i=1}^{n-p-q-1} E(x_{q_i} x_{p, i+q+1}) + \sum_{i=1}^{n-p-q-2} E(x_{q_i} x_{p, i+q+2}) \\ &= [(n-p-q-2)E(x_{q_i} x_{p, i+q}) + (n-p-q-3)E(x_{q_i} x_{p, i+q+1}) \\ &\quad + (n-p-q-4)E(x_{q_i} x_{p, i+q+2}) \\ &\quad + E(x_{q_1} x_{p, q+1}) + E(x_{q_1} x_{p, q+2}) + E(x_{q_1} x_{p, q+3}) \\ &\quad + E(x_{q, n-q-p} x_{p, n-p}) + E(x_{q, n-q-p-1} x_{p, n-p}) + E(x_{q, n-q-p-2} x_{p, n-p})]. \end{aligned}$$

Now $E(x_{q_i} x_{p, i+q}) = 2P|- , +^q, -^p, +| = 2A$, where

$$\begin{aligned} A &= \int_0^1 \int_0^{y_{i+p+q+1}} \int_{y_{i+p+q}}^1 \cdots \int_{y_{i+q+1}}^1 \left[\int_0^{y_{i+q}} \cdots \int_0^{y_{i+1}} \int_{y_i}^1 dy_{i-1} \cdots dy_{i+q-1} \right] \\ &\quad \cdot dy_{i+q} \cdots dy_{i+p+q+1}. \end{aligned}$$

The expression within the square brackets is easily evaluated, and applying (3.6) to the result, we have

$$A = \frac{1}{(p+q+1)p!q!} - \frac{1}{(p+q+2)(p+1)!q!} \\ - \frac{1}{(p+q+2)p!(q+1)!} + \frac{1}{(p+q+3)(p+1)!(q+1)!}.$$

Similarly, $E(x_q, x_{p, i+q+1}) = 2P\{-, +^q, -, +^p, -\}$, and $E(x_q, x_{p, i+q+2}) = 2P\{-, +^q, -, -, +^p, -\} + 2P\{-, +^q, -, +, -^p, +\}$. The other terms in the right member of (4.7) are obtained in like manner. The right member of (4.7) is symmetric in p and q ; hence the second and third terms of the right member of (4.5) are equal.

We now consider the remaining terms in the right member of (4.5) for $p > q$; the result obtained also holds for $p \leq q$. We write them as

$$(4.8) \quad - \left\{ \sum_{i=p+3}^{n-q} E(x_q)E(x_{p, i-(p+2)}) + \sum_{i=p+2}^{n-q} E(x_q)E(x_{p, i-(p+1)}) \right. \\ + \cdots + \sum_{i=p-q+1}^{n-q} E(x_q)E(x_{p, i-(p-q)}) + \sum_{i=p-q}^{n-q-1} E(x_q)E(x_{p, i-(p-q-1)}) \\ + \cdots + \sum_{i=1}^{n-p} E(x_q)E(x_{p, i}) + \sum_{i=1}^{n-p-1} E(x_q)E(x_{p, i+1}) \\ \left. + \cdots + \sum_{i=1}^{n-p-(q+2)} E(x_q)E(x_{p, i+(q+2)}) \right\}.$$

The $(p+q+5)$ sums in (4.8) comprise in all $\left\{ (n-p)(p+q+5) - 2 \sum_{k=1}^{q+2} k \right\}$ terms. Remembering that $E(x_{p, n-p}) = E(x_{p1})$, (4.8) becomes

$$(4.9) \quad - \{ [n(p+q+5) - (p^2 + pq + q^2 + 7p + 7q + 16)] E(x_q)E(x_{p1}) \\ + [2p+4] E(x_q)E(x_{p1}) + [2q+4] E(x_{q1})E(x_{p1}) + 2E(x_{q1})E(x_{p1}) \}.$$

Adding the right member of (4.6), twice the right member of (4.7), and (4.9), we have

$$\sigma(r_p r_q) = \\ 2n \left\{ -2 \frac{p^2(p+q+6)(q^2+3q+1) + p(3q^3+20q^2+40q+19) + (q^3+9q^2+29q+26)}{(q+3)!(p+3)!} \right. \\ + 2 \frac{-p+q+1}{(p+q+3)(q+2)!(p+1)!} - 2 \frac{1}{(p+q+5)(q+3)!(p+1)!} \\ - 2 \frac{(p+q)^3 + 9(p+q)^2 + 23(p+q) + 14}{(p+q+5)!} \\ \left. + 2 \frac{1}{(p+q+1)q!p!} + \delta_{pq} \frac{p^2+3p+1}{(p+3)!} \right\}$$

$$\begin{aligned}
(4.10) \quad & + 2 \frac{2}{(q+3)!(p+3)!} \left[\begin{aligned} & p^4(q^2+3q+1) + p^3(q^3+9q^2+19q+6) \\ & + p^2(q^4+9q^3+28q^2+35q+11) \\ & + p(3q^4+20q^3+40q^2+29q+10) \\ & + (q^4+9q^3+27q^2+32q+10) \end{aligned} \right] \\
& + 2 \frac{(p+q+2)(p-q-1)}{(p+q+3)(q+2)!(p+1)!} + 2 \frac{p+q+4}{(p+q+5)(q+3)!(p+1)!} \\
& + 2 \frac{(p+q)^4+10(p+q)^3+29(p+q)^2+16(p+q)-19}{(p+q+5)!} \\
& - 2 \left\{ \frac{p+q}{(p+q+1)q!p!} - \delta_{pq} \frac{p^3+3p^2-p-4}{(p+3)!} \right\},
\end{aligned}$$

where δ_{pq} is the Kronecker delta. Formula (4.10) is valid for $p+q \leq n-4$. It is symmetric in p and q . Setting $p=q$ we obtain

$$\begin{aligned}
(4.11) \quad \sigma^2(r_p) = 2n \left\{ -2 \frac{2p^5+15p^4+41p^3+55p^2+48p+20}{(p+3)!(p+3)!} \right. \\
+ 2 \frac{2p^3+9p+12}{(2p+3)(2p+5)(p+3)!(p+1)!} - 4 \frac{4p^3+18p^2+23p+7}{(2p+5)!} \\
+ 2 \left\{ \frac{1}{(2p+1)p!p!} + \frac{p^2+3p+1}{(p+3)!} \right\} \\
+ 2 \left\{ 2 \frac{3p^5+24p^4+69p^3+90p^2+67p^2+42p+10}{(p+3)!(p+3)!} \right. \\
- 4 \frac{2p^3+11p^2+19p+9}{(2p+3)(2p+5)(p+3)!(p+1)!} \\
+ 2 \frac{16p^4+80p^3+110p^2+32p-19}{(2p+5)!} \\
\left. - 4 \left\{ \frac{p}{(2p+1)p!p!} - \frac{p^3+3p^2-p-4}{(p+3)!} \right\} \right\}
\end{aligned}$$

We next evaluate $\sigma(r'_p r'_q)$. Since w_{qi} and w_{pj} are independent for $j \geq i+q+2$, we have, corresponding to (4.5),

$$\begin{aligned}
(4.12) \quad \sigma(r'_p r'_q) = \sum_i E(w_{pi} w_{qi}) + \sum_{i < j < i+q+2} E(w_{qi} w_{pj}) + \sum_{i < j < i+p+2} E(w_{pi} w_{qj}) \\
- \left[\sum_i E(w_{pi}) E(w_{qi}) + \sum_{i < j < i+q+2} E(w_{qi}) E(w_{pj}) \right. \\
\left. + \sum_{i < j < i+p+2} E(w_{pi}) E(w_{qj}) \right].
\end{aligned}$$

Let $G = \text{Max}(p, q)$. Then $w_p, w_q \equiv w_G$ and we have for the first term of the right member of (4.12)

$$(4.13) \quad \sum_{i=1}^n E(w_p, w_q) = E(r'_G).$$

The second term in the right member of (4.12) may be written

$$(4.14) \quad (n-p-q-1)E(w_q, w_{p,1+q}) + (n-p-q-2)E(w_q, w_{p,1+q+1}) \\ + E(w_q, w_{p,q+1}) + E(w_q, w_{p,q+2}).$$

Now $E(w_q, w_{p,1+q}) = 2P\{-, +^q, -^p\}$, $E(w_q, w_{p,1+q+1}) = 2P\{-, +^q, -, +^p\} + 2P\{-, +^{q+1}, -^p\}$, and the other terms are obtained similarly. The third term in the right member of (4.12) will be equal to (4.14) with p and q interchanged.

The remaining terms in the right member of (4.12) become

$$(4.15) \quad -\{[n(p+q+3) - (p^2 + pq + q^2 + 4p + 4q + 5)]E(w_q)E(w_p) \\ + [p+1]E(w_q)E(w_{p1}) + [q+1]E(w_{q1})E(w_p) + E(w_{q1})E(w_{p1})\}.$$

We can now write the formula for $\sigma(r'_p, r'_q)$, valid for $p+q \leq n-2$,

$$(4.16) \quad \sigma(r'_p, r'_q) = 2n \left\{ -\frac{p^2(2q+2) + p(2q^2 + 8q + 5) + (2q+1)(q+2)}{(q+2)!(p+2)!} \right. \\ + \frac{2}{(p+q+1)q!p!} - \frac{(q+1)(q+2) + (p+1)(p+2)}{(p+q+3)(q+2)!(p+2)!} \\ \left. - 2 \frac{p+q+2}{(p+q+3)!} + \frac{(G+1)}{(G+2)!} \right\} \\ + 2 \left\{ \frac{p^3(2q+2) + p^2(2q^2 + 8q + 5) + p(2q^3 + 8q^2 + 6q - 2) + (2q^3 + 5q^2 - 2q - 6)}{(q+2)!(p+2)!} \right. \\ - 2 \frac{p+q}{(p+q+1)q!p!} + \frac{(p+q+2)[(p+1)(p+2) + (q+1)(q+2)]}{(p+q+3)(q+2)!(p+2)!} \\ \left. + 2 \frac{(p+q)^2 + 3(p+q) + 1}{(p+q+3)!} - \frac{G^2 + G - 1}{(G+2)!} \right\},$$

where $G = \text{Max}(p, q)$. Setting $p = q$ we obtain

$$(4.17) \quad \sigma^2(r'_p) = 2n \left\{ -2 \frac{(p+1)(2p^2 + 4p + 1)}{(p+2)!(p+2)!} + 2 \frac{1}{(2p+1)p!p!} \right. \\ - 2 \frac{1}{(2p+3)(p+2)!p!} - 4 \frac{p+1}{(2p+3)!} + \frac{p+1}{(p+2)!} \left. \right\} \\ + 2 \left\{ 2 \frac{(p+1)^2(3p^2 + 4p - 3)}{(p+2)!(p+2)!} - 4 \frac{p}{(2p+1)p!p!} \right. \\ \left. + 4 \frac{p+1}{(2p+3)(p+2)!p!} + 2 \frac{4p^2 + 6p + 1}{(2p+3)!} - \frac{p^2 + p - 1}{(p+2)!} \right\}.$$

Setting $p = 1$, we have

$$(4.18) \quad \sigma^2(r'_1) = \sigma^2(r) = \frac{16n - 20}{90}.$$

The value of $\sigma^2(r)$ was given by Bienaymé [2].

Finally, we evaluate

$$(4.19) \quad \begin{aligned} \sigma(r_p r'_q) &= \sum_i E(x_{p_i} w_{q_i}) + \sum_{i < j < i+q+2} E(w_{q_i} x_{p_j}) + \sum_{i < j < i+p+3} E(x_{p_i} w_{q_j}) \\ &= [\sum_i E(x_{p_i}) E(w_{q_i}) + \sum_{i < j < i+q+2} E(w_{q_i}) E(x_{p_j}) \\ &\quad + \sum_{i < j < i+p+3} E(x_{p_i}) E(w_{q_j})]. \end{aligned}$$

Let the symbol $\eta_{pq} = \begin{cases} 1, & i \geq j \\ 0, & i < j \end{cases}$. Then $x_{p_i} w_{q_i} = \eta_{pq} x_{p_i}$, and

$$(4.20) \quad \sum_{i=1}^n E(x_{p_i} w_{q_i}) = \eta_{pq} [E(r_p)].$$

The remaining terms of (4.19) introduce no new difficulties, and for $p + q \leq n - 3$ we obtain

$$(4.21) \quad \begin{aligned} \sigma(r_p r'_q) &= 2n \left\{ - \frac{p^2(2q+2) + p^2(2q^2+13q+12)}{(p+3)!(q+2)!} \right. \\ &\quad + \frac{2}{(p+q+1)p!q!} + \frac{p-q}{(p+q+2)(p+1)!(q+1)!} \\ &\quad + \frac{(p-q+1)(q+2)}{(p+q+3)(p+2)!(q+2)!} + \frac{(p+2)(p+3) + (q+1)(q+2)}{(p+q+4)(p+3)!(q+2)!} \\ &\quad \left. - 2 \frac{(p+q)^2 + 5(p+q) + 5}{(p+q+4)!} + \eta_{pq} \frac{p^2 + 3p + 1}{(p+3)!} \right\} \\ &\quad + 2 \left\{ \frac{1}{(p+3)!(q+2)!} \left[\begin{aligned} &p^4(2q+2) + p^3(2q^2+13q+12) \\ &+ p^2(2q^3+13q^2+26q+24) \\ &+ p(6q^3+22q^2+19q+27) \\ &+ (2q^3+6q^2+10q+25) \end{aligned} \right] \right. \\ &\quad - 2 \frac{p+q}{(p+q+1)p!q!} - \frac{p^2+2p-q^2+2}{(p+q+2)(p+1)!(q+1)!} \\ &\quad - \frac{(p+2)[(p+2)(q+3)-1] - q(q+1)(q+2)}{(p+q+3)(p+2)!(q+2)!} \\ &\quad - \frac{(p+q+3)[(p+2)(p+3) + (q+1)(q+2)]}{(p+q+4)(p+3)!(q+2)!} \\ &\quad \left. + 2 \frac{(p+q)^3 + 6(p+q)^2 + 8(p+q) - 1}{(p+q+4)!} - \eta_{pq} \frac{p^3 + 3p^2 - p - 4}{(p+3)!} \right\}, \end{aligned}$$

where η_{pq} is defined as in (4.20).

5. The use of runs up and down in tests of significance. Certain misconceptions about the application of runs up and down have appeared in the literature, and it is the purpose of this section to clarify them.

Since $E(r_p)$, $\sigma^2(r_p)$, $E(r)$ and $\sigma^2(r)$ are all of the order n , it follows that r_p/r converges stochastically to

$$\lambda_p = \lim_{n \rightarrow \infty} \frac{E(r_p)}{E(r)}.$$

Let

$$\lambda'_p = \lim_{n \rightarrow \infty} \frac{E(r'_p)}{E(r)}.$$

From (4.1) and (4.2) we have

$$\lambda_1 = \frac{5}{8} = .6250$$

$$\lambda_2 = \frac{11}{40} = .2750$$

$$\lambda_3 = \frac{19}{240} = .07917$$

$$\lambda_4 = \frac{29}{1680} = .01726$$

$$\lambda'_5 = \frac{1}{280} = .00357$$

Let

$$\lambda_{pn} = \frac{E(r_p)}{E(r)}.$$

Some writers say that λ_{pn} or λ_p is "the probability of a run of length p ." If the stochastic process consists in obtaining a sequence from among the $n!$ sequences S , each of which has the probability $(n!)^{-1}$, then the phrase "the probability of a run of length p " has no meaning. One can speak of the probability of at least one run of length p (i.e., that $r_p > 0$), of the probability of no run of length p ($r_p = 0$), of the probability that the first or fifth run (if there are five runs) in the sequence S be of length p , etc. It is possible to give *different* stochastic processes in which "the probability of a run of length p " will have meaning and be λ_{pn} , or λ_p . Consider, for example, the totality of all the runs in the $n!$ sequences S . There are $n!E(r)$ of them, and among these there are $n!E(r_p)$ runs of length p . Now let the stochastic process consist in drawing a run from the totality of all these runs, each of which is to have the same probability, which is therefore $[n!E(r)]^{-1}$. Then the probability of drawing a run of length p is λ_{pn} . It is difficult to see how this stochastic process can have rele-

vance to most of the problems of quality control and economic time series where runs up and down are now used.

Some writers on quality control and economic time series recommend that statistical control or randomness be tested by use of $d_1, \dots, d_{p-1}, d'_p$, where

$$d_i = r_i - E(r_i), \quad (i = 1, 2, \dots, (p-1)),$$

$$d'_p = r'_p - E(r'_p).$$

The availability of the covariance matrix M of $d_1, \dots, d_{p-1}, d'_p$, which we have obtained in this paper, will assist in the construction of such tests. Also of help will be a result recently announced by one of us [7], the early publication of which is expected. This result states that in the limit with n the joint probability density function of $d_1, \dots, d_{p-1}, d'_p$, is $K e^{-Q}$, where K is a constant and Q is a quadratic form in $d_1, \dots, d_{p-1}, d'_p$ whose matrix is the inverse of the matrix M . It follows immediately that Q has in the limit the χ^2 distribution with p degrees of freedom.

We wish now to make a few remarks about the tests of significance, based on runs up and down, which are used by some contemporary writers. A description of their method can perhaps be best given by an example. With $n = 100$ and $p = 5$, say, suppose the observed values are:

Observed Values

$$\begin{array}{rcl} r_1 & = & 30 \\ r_2 & = & 10 \\ r_3 & = & 4 \\ r_4 & = & 3 \\ r'_5 & = & 3 \\ \hline \text{Total, } r & = & 50 \end{array}$$

These writers then say that the expected values are:

Expected Values according to some writers

$$\begin{array}{rcl} E(r_1) = r\lambda_1 = 50 (.6250) & = & 31.25 \\ E(r_2) = r\lambda_2 = 50 (.2750) & = & 13.75 \\ E(r_3) = r\lambda_3 = 50 (.07917) & = & 3.96 \\ E(r_4) = r\lambda_4 = 50 (.01726) & = & 0.86 \\ E(r'_5) = r\lambda'_5 = 50 (.00357) & = & 0.18 \\ & & \hline & & 50.00 \end{array}$$

The correct expected values are given by (4.1) and (4.2) and are:

Correct Expected Values

$$\begin{array}{rcl} E(r_1) & = & 41.75 \\ E(r_2) & = & 18.10 \\ E(r_3) & = & 5.15 \\ E(r_4) & = & 1.11 \\ E(r'_5) & = & 0.22 \\ & & \hline & & 66.33 \end{array}$$

It should be noted that:

(a) A consequence of the incorrect method of obtaining "expected values" is that, since

$$E(r) = E(r_1) + E(r_2) + E(r_3) + E(r_4) + E(r'_5),$$

it implies that the *expected* number of runs of all lengths is equal to the *observed* number! This is obviously erroneous. In fact it follows from (4.18) and the results announced in [7] that $r - E(r)$ is in the probability sense of order \sqrt{n} .

(b) By using the incorrect expected values for comparison with the observed values one loses the valuable information furnished by $r - E(r)$. If this is large (in terms of its standard deviation) it is plausible to question whether statistical control or randomness exists.

6. Summary. Let $S = (h_1, \dots, h_n)$ be a random permutation of the n unequal numbers a_1, \dots, a_n , and let R be the sequence of signs (+ or -) of the differences $h_{i+1} - h_i$ ($i = 1, \dots, n-1$). It is assumed that each of the $n!$ sequences S is equally probable. A sequence of p successive + (-) signs not immediately preceded or followed by a + (-) sign is called a run up (down) of length p . Let r_p and r'_p be the number of runs up and down in R of lengths p and p or more respectively. In this paper the exact values of $\sigma(r_p r_q)$, (see formula (4.10)); $\sigma^2(r_p)$, (formula (4.11)); $\sigma(r'_p r'_q)$, (formula (4.16)); $\sigma^2(r'_p)$, (formula (4.17)); and $\sigma(r_p r'_q)$, (formula (4.21)) are derived. A few numerical values are:

$$\begin{aligned} \sigma^2(r_1) &= \frac{305n - 347}{720}, & \sigma^2(r_2) &= \frac{51106n - 73859}{453600}, \\ \sigma^2(r'_1) &= \frac{16n - 29}{90}, & \sigma^2(r'_2) &= \frac{57n - 43}{720}, & \sigma^2(r'_3) &= \frac{21496n - 51269}{453600}, \\ \sigma(r_1 r_2) &= -\frac{19n + 11}{210}, & \sigma(r'_1 r'_2) &= -\frac{5n - 3}{60}, & \sigma(r'_1 r'_3) &= \\ & -\frac{41n - 99}{630}, & \sigma(r_2 r'_1) &= -\frac{23n + 135}{1260}, & \sigma(r_1 r'_2) &= -\frac{117n - 79}{720}, \\ \sigma(r_1 r'_3) &= -\frac{363n - 817}{5040}, & \text{and } \sigma(r_2 r'_3) &= -\frac{18346n - 49019}{453600}. \end{aligned}$$

The values of $E(r_p)$, (formula (4.1)); and $E(r'_p)$, (formula (4.2)) are also given. Certain misconceptions about the applications of runs up and down are discussed.

REFERENCES

- [1] L. BESSON, (transl. and abr. by E. W. Woolard), "On the comparison of meteorological data with chance results," *Monthly Weather Review*, (U. S. Weather Bureau), Vol. 48 (1920), pp. 89-94.
- [2] J. BRENAYMÉ, "Sur une question de probabilités," *Bull. Soc. Math. France*, Vol. 2 (1874), pp. 153-154.

- [3] R. A. Fisher, "On the random sequence," *Quarterly Jour. Roy. Meteorological Soc.*, Vol. 52 (1926), p. 250.
- [4] W. G. KENNEDY and A. G. McKENDRICK, "Tests for randomness in a series of observations," *Proc. Roy. Soc. Edinburgh*, Vol. 57 (1937), pp. 228-240.
- [5] W. G. KENNEDY and A. G. McKENDRICK, "Some distributions associated with a randomly arranged set of numbers," *Proc. Roy. Soc. Edinburgh*, Vol. 57 (1937), pp. 332-376.
- [6] W. A. WALLIS and L. H. MOORE, *A significance test for time series*, Technical Paper 1, Nat. Bureau of Econ. Res., New York, 1941.
- [7] J. WILKOWITZ, "Asymptotic distributions of ascending and descending runs," (abstract) *Bull. Amer. Math. Soc.*, Vol. 49 (1943), pp. 539-540.

ON THE MEASURE OF A RANDOM SET

BY H. E. ROBBINS

Post Graduate School, U. S. Naval Academy

1. Introduction. The following is perhaps the simplest non-trivial example of the type of problem to be considered in this paper. On the real number axis let N points x_i ($i = 1, 2, \dots, N$) be chosen independently and by the same random process, so that the probability that x_i shall lie to the left of any point x is a given function of x ,

$$(1) \quad \sigma(x) = \Pr(x_i < x).$$

With the points x_i as centers, N unit intervals are drawn. Let X denote the set-theoretical sum of the N intervals, and let $\mu(X)$ denote the linear measure of X . Then $\mu(X)$ will be a chance variable whose values may range from 1 to N , and whose probability distribution is completely determined by $\sigma(x)$. Let $\tau(u)$ denote the probability that $\mu(X)$ be less than u . Then by definition, the expected value of $\mu(X)$ is

$$(2) \quad E(\mu(X)) = \int_1^N u \, d\tau(u),$$

where

$$(3) \quad \tau(u) = \Pr(\mu(X) < u).$$

The problem is to transform the expression for $E(\mu(X))$ so that its value may be computed in terms of the given function $\sigma(x)$.

In order to do this, we observe that, since the x_i are independent,

$$(4) \quad \tau(u) = \int \cdots \int_{C(u)} d\sigma(x_1) \cdots d\sigma(x_N),$$

where the domain of integration $C(u)$ consists of all points (x_1, \dots, x_N) in Euclidean N -dimensional space such that the linear measure of the set-theoretical sum of N unit intervals with centers at the points x_i is less than u . Here, however, a difficulty arises. Due to the possible overlapping of the intervals, the geometrical description of the domain $C(u)$ is such as to make the explicit evaluation of the integral (4) a complicated matter.

The difficulty is even more serious in the analogous problem where instead of N unit intervals on the line we have N unit circles in the plane, with a given probability distribution for their centers (x_i, y_i) . Again we seek the expected value of the measure of the set-theoretical sum of the N circles. The corresponding domain $C(u)$ in $2N$ -dimensional space will now be very complicated.

It is the object of this paper to show how, in such cases as these, the expected value of $\mu(X)$ may be found without first finding the distribution function $\tau(u)$.

In fact, the theorem to be stated in (15) will in many important cases yield a comparatively simple formula for $E(\mu(X))$.

2. Expected value of $\mu(X)$. In order to state the problem in full generality, let us suppose that X is a random Lebesgue measurable subset of Euclidean n dimensional space E_n . By this we shall mean that in the space T of all possible values of X there is defined a probability measure $\rho(X)$ so that for every ρ -measurable subset S of T , the probability that X shall belong to S is given by the Lebesgue-Stieltjes integral

$$(5) \quad \Pr \{X \in S\} = \int_T C_S(X) d\rho(X),$$

where the integrand is the characteristic function of S ,

$$(6) \quad C_S(X) = \begin{cases} 1 & \text{for } X \in S \\ 0 & \text{for } X \notin S. \end{cases}$$

In practice, the set X will be a function of a finite number of real parameters (e.g., the coordinates of the centers of the intervals or circles considered in the Introduction), $X = X(\alpha_1, \dots, \alpha_r) = X(\alpha)$. There will be given a probability measure $\nu(\alpha)$ in the parameter space E_r , so that α will be a vector random variable in the ordinary sense. If A is any ν -measurable subset of E_r , then by definition,

$$(7) \quad \Pr \{\alpha \in A\} = \int_A C_A(\alpha) d\nu(\alpha).$$

Now for the set S' consisting of all X such that $X = X(\alpha)$ for α in A , we define $\rho(S') = \nu(A)$. Thus a ρ -measure is defined in the space T of X , which is the general situation considered in the preceding paragraph.

Returning to the general case described in the first paragraph of this section, we shall now prove the main theorem of this paper. To this end we define, for every point x of E_n and every set X of T , the function

$$(8) \quad g(x, X) = \begin{cases} 1 & \text{for } x \in X \\ 0 & \text{for } x \notin X. \end{cases}$$

Moreover, for every x in E_n we let $S(x)$ denote the set of all X in T which contain x . Then for every x in E_n we have from (6),

$$(9) \quad g(x, X) = C_{S(x)}(X).$$

Let us denote the Lebesgue measure in E_n of the set X by $\mu(X)$. Assuming that the function $g(x, X)$ is a $\mu\rho$ -measurable function of the pair (x, X) in the product space¹ of E_n with T , it follows from Fubini's theorem¹ that

$$(10) \quad \int_{E_n \times T} g(x, X) d\mu\rho(x, X) = \int_{E_n} \int_T g(x, X) d\rho(X) d\mu(x).$$

¹ See S. Saks, *Theory of the Integral*, G. E. Stechert, N. Y., 1937, pp. 86, 87.

From (5) and (9) it follows that

$$(11) \quad \int_T g(x, X) d\rho(X) = \Pr(X \in S(x)) = \Pr(x \in X).$$

Again by Fubini's theorem we have

$$(12) \quad \int_{E_n \times T} g(x, X) d\mu(x, X) = \int_T \int_{E_n} g(x, X) d\mu(x) d\rho(X)$$

But from (8),

$$(13) \quad \int_{E_n} g(x, X) d\mu(x) = \int_X d\mu(x) = \mu(X).$$

Now from (10), (11), (12), and (13) we have

$$(14) \quad \int_{E_n} \Pr(x \in X) d\mu(x) = \int_T \mu(X) d\rho(X).$$

But the latter integral is equal to $E(\mu(X))$. Hence we have the relation

$$(15) \quad E(\mu(X)) = \int_{E_n} \Pr(x \in X) d\mu(x).$$

This is our fundamental result. We may state it as a

THEOREM: Let X be a random Lebesgue measurable subset of E_n , with measure $\mu(X)$. For any point x of E_n let $p(x) = \Pr(x \in X)$. Then, assuming that the function $g(x, X)$ defined by (8) is a measurable function of the pair (x, X) , the expected value of the measure of X will be given by the Lebesgue integral of the function $p(x)$ over E_n .

3. Higher moments of $\mu(X)$. We may generalize the result (15) to obtain similar expressions for the higher moments of $\mu(X)$. For the second moment we have the expression

$$(16) \quad E(\mu^2(X)) = \int_T \mu^2(X) d\rho(X).$$

Now from (13),

$$(17) \quad \begin{aligned} \mu^2(X) &= \mu(X) \cdot \mu(X) = \int_{E_n} g(x, X) d\mu(x) \cdot \int_{E_n} g(y, X) d\mu(y) \\ &= \int_{E_n} \int_{E_n} g(x, X) \cdot g(y, X) d\mu(x) d\mu(y). \end{aligned}$$

Let

$$(18) \quad g(x, y, X) = g(x, X) \cdot g(y, X) = \begin{cases} 1 & \text{if } X \text{ contains both } x \text{ and } y \\ 0 & \text{otherwise.} \end{cases}$$

Then from (16), (17), and (18), we have as before by Fubini's theorem,

$$(19) \quad \begin{aligned} E(\mu^2(X)) &= \int_T \int_{E_n} \int_{E_n} g(x, y, X) d\mu(x) d\mu(y) d\rho(X) \\ &= \int_{E_n} \int_{E_n} \int_T g(x, y, X) d\rho(X) d\mu(x) d\mu(y). \end{aligned}$$

But from (5) and (18) it follows that

$$(20) \quad \int_T g(x, y, X) d\rho(X) = \Pr(x \in X \text{ and } y \in X).$$

The latter probability may be denoted by $p(x, y)$. This function will be defined over the Cartesian product, E_{2n} , of E_n with itself. Let $\mu(x, y)$ denote Lebesgue measure in E_{2n} . Then from (19) we have

$$(21) \quad E(\mu^2(X)) = \int_{E_{2n}} p(x, y) d\mu(x, y),$$

where

$$(22) \quad p(x, y) = \Pr(x \in X \text{ and } y \in X).$$

The formula for the m th moment of $\mu(X)$ will clearly be

$$(23) \quad \text{Exp}(\mu^m(X)) = \int_{E_{mn}} p(x_1, x_2, \dots, x_m) d\mu(x_1, x_2, \dots, x_m),$$

where $\mu(x_1, x_2, \dots, x_m)$ denotes Lebesgue measure in E_{mn} and where

$$(24) \quad p(x_1, x_2, \dots, x_m) = \Pr(x_1 \in X \text{ and } x_2 \in X \dots \text{ and } x_m \in X).$$

In the next section we shall apply formulas (15) and (21) to a specific problem.

4. Let a, p, B be given positive numbers such that $(B + a)p \leq a$ and $a \leq B$. We shall define the random linear point set X as follows. N intervals, each of length a , are chosen independently on the number axis. The probability density function for the center of the i th interval will be assumed to be constant and equal to p/a in the interval $-a/2 \leq x \leq B + (a/2)$; it may be arbitrary outside this interval. The set X is now defined as the intersection of the fixed interval $I: 0 \leq x \leq B$ with the variable set-theoretical sum of the N intervals. The hypothesis of (15) is clearly satisfied. The probability that any point x in the interval I shall be contained in the i th interval of length a is clearly $(p/a)a = p$. From this it follows that

$$(25) \quad \Pr(x \in X) = p(x) = \begin{cases} 1 - (1 - p)^N & \text{for } 0 \leq x \leq B \\ 0 & \text{elsewhere.} \end{cases}$$

From (15) it follows that

$$(26) \quad E(\mu(X)) = \int_0^B p(x) dx = B(1 - (1 - p)^N).$$

(The same formula holds in the case where the N intervals of length a are replaced by N circles of area a and I by a plane domain of area B , provided that for every point of the domain the probability of being contained in the i th circle is equal to a constant p . A similar remark holds for spheres in space.)

To evaluate $E(\mu^2(X))$ in the linear case we make use of the identity

$$(27) \quad \Pr(A \text{ and } B) = \Pr(A) + \Pr(B) + \Pr(\text{neither } A \text{ nor } B) - 1,$$

which holds for any two events A and B . It follows from (27) and (25) that if x and y are any two points of I , then

$$\begin{aligned} p(x, y) &= \Pr(x \in X \text{ and } y \in X) \\ (28) \quad &= \Pr(x \in X) + \Pr(y \in X) + \Pr(x \notin X \text{ and } y \notin X) - 1 \\ &= 1 - 2(1 - p)^N + \Pr(x \notin X \text{ and } y \notin X). \end{aligned}$$

Let

$$(29) \quad h(x, y) = \Pr(x \notin X \text{ and } y \notin X).$$

Then

$$(30) \quad h(x, y) = \begin{cases} [1 - (p/a)2a]^N = (1 - 2p)^N, & \text{for } |y - x| \geq a \\ [1 - (p/a)(a + |y - x|)]^N = \left(\frac{a - ap - p|y - x|}{a}\right)^N, & \text{for } |y - x| < a. \end{cases}$$

Now from (21), (28), and (29) we have

$$\begin{aligned} E(\mu^2(X)) &= \int_0^B \int_0^B p(x, y) dy dx \\ (31) \quad &= \int_0^B \int_0^B [1 - 2(1 - p)^N + h(x, y)] dy dx \\ &= B^2[1 - 2(1 - p)^N] + 2 \int_0^B \int_x^B h(x, y) dy dx. \end{aligned}$$

When the latter integral is evaluated the result is

$$\begin{aligned} E(\mu^2(X)) &= B^2[1 - 2(1 - p)^N] + (B - a)^2(1 - 2p)^N \\ (32) \quad &+ \frac{2aB(1 - p)^{N+1}}{(N + 1)p} - \frac{2a(B - a)(1 - 2p)^{N+1}}{(N + 1)p} \\ &- \frac{2a^2}{(N + 1)(N + 2)p^2} [(1 - p)^{N+2} - (1 - 2p)^{N+2}]. \end{aligned}$$

Combining this with (26), we find for the variance of $\mu(X)$ the expression

$$\begin{aligned} \sigma^2 &= E(\mu^2(X)) - [E(\mu(X))]^2 \\ (33) \quad &= (B - a)^2(1 - 2p)^N - B^2(1 - p)^{2N} + \frac{2aB(1 - p)^{N+1}}{(N + 1)p} \\ &- \frac{2a(B - a)(1 - 2p)^{N+1}}{(N + 1)p} - \frac{2a^2}{(N + 1)(N + 2)p^2} [(1 - p)^{N+2} - (1 - 2p)^{N+2}]. \end{aligned}$$

ON THE DISTRIBUTION OF THE RADIAL STANDARD DEVIATION

BY FRANK E. GRUBBS¹

Aberdeen Proving Ground

1. Introduction. Of interest in the field of ballistics is a measure of the accuracy of bullets. In acceptance tests of small arms ammunition lots, for example, a sample of rounds from each lot is fired from a fixed rifle at a vertical target placed a specified distance from the rifle. The accuracy of the bullets is taken to be some measure of the scattering (or lack of scattering) of the bullet holes on the target. The purpose of such a test would be to determine whether or not the lot under consideration differs significantly in accuracy from (a) standard values or (b) its predecessors.

One useful measure of accuracy is the radial standard deviation which is defined by the relation

$$(1) \quad Z = \sqrt{\frac{1}{N} \{ \sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2 \}},$$

where x_i and y_i are respectively the abscissa and ordinate of any point measured from an arbitrary origin and N is the sample size.

It will be the purpose of the present discussion to call attention to a series expansion for the distribution of the statistic Z in samples of N assuming that the distribution of all rounds of the lot on the target follow the bivariate normal population law

$$(2) \quad f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{x^2}{2\sigma_1^2} - \frac{y^2}{2\sigma_2^2}}, \quad (x \text{ and } y \text{ statistically independent})$$

where σ_1^2 and σ_2^2 are the parent variances of x and y respectively. In the above probability density function, the population means are taken to be zero since the statistic Z is quite independent of the origin selected.

2. Moment generating function of Z^2 . The distribution of $s_1^2 = \frac{1}{N} \sum (x_i - \bar{x})^2$ in samples of N from a normal population is given by the well-known law,

$$(3) \quad dF(s_1^2) = \frac{\frac{N}{2\sigma_1^2}}{\Gamma\left(\frac{N-1}{2}\right)} \left(\frac{Ns_1^2}{2\sigma_1^2}\right)^{\frac{1}{2}(N-2)} e^{-\frac{Ns_1^2}{2\sigma_1^2}} ds_1^2, \quad s_1^2 \geq 0.$$

The moment generating function of s_1^2 may be found (in a neighborhood of $t = 0$) by straightforward integration:

$$(4) \quad M_{s_1^2}(t) = E(e^{t s_1^2}) = \int_0^\infty e^{t s_1^2} dF(s_1^2) = \left\{ 1 - \frac{2\sigma_1^2 t}{N} \right\}^{-\frac{1}{2}(N-1)}.$$

¹ Captain, Ordnance Department, Ballistic Research Laboratory, Aberdeen Proving Ground, Md.

Likewise, for $s_2^2 = \frac{1}{N} \sum (y_i - \bar{y})^2$, we have

$$(5) \quad M_{s_2^2}(t) = \left\{ 1 - \frac{2\sigma_2^2 t}{N} \right\}^{-1(N-1)}$$

Now $M_{z^2}(t) = M_{s_1^2+s_2^2}(t) = E\{e^{it(s_1^2+s_2^2)}\} = E(e^{it s_1^2}) \cdot E(e^{it s_2^2})$ since x and y are independent. Thus,

$$M_{z^2}(t) = M_{s_1^2}(t) \cdot M_{s_2^2}(t) = \left\{ 1 - \frac{2\sigma_1^2 t}{N} \right\}^{-1(N-1)} \left\{ 1 - \frac{2\sigma_2^2 t}{N} \right\}^{-1(N-1)}$$

3. Distribution function of Z^2 . Making use of the Fourier theorem, we have

$$(6) \quad f(Z^2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left\{ 1 - \frac{2\sigma_1^2 it}{N} \right\}^{-1(N-1)} \left\{ 1 - \frac{2\sigma_2^2 it}{N} \right\}^{-1(N-1)} e^{-iz^2 t} dt,$$

at all points of continuity of $f(Z^2)$.

The discussion will be divided preferably into the two cases: Case I: $\sigma_1^2 = \sigma_2^2$, and Case II: $\sigma_1^2 \neq \sigma_2^2$.

Case I: $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

In this case the distribution of Z^2 reduces to

$$(7) \quad f(Z^2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left\{ 1 - \frac{2\sigma^2 it}{N} \right\}^{-(N-1)} e^{-iz^2 t} dt.$$

It will simplify the algebra to find first the distribution of $u^2 = \frac{NZ^2}{2\sigma^2}$ and then that of Z^2 . Since $M_{u^2}(t) = \{1 - it\}^{-(N-1)}$,

$$(8) \quad f(u^2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \{1 - it\}^{-(N-1)} e^{-iu^2 t} dt.$$

This integral may be evaluated easily by the calculus of residues since the integrand has only a single pole of order $(N-1)$ at $t = -i$. We will, however, make use of the following method.

Put $-v = u^2 - iu^2 t$; then

$$(9) \quad \begin{aligned} f(u^2) &= \frac{1}{2\pi} \int_{u^2-i\infty}^{-u^2+i\infty} \left(-\frac{v}{u^2} \right)^{-(N-1)} e^{-v-u^2} \frac{dv}{iu^2} \\ &= \frac{-e^{-u^2} (u^2)^{N-2}}{2\pi i} \int_{u^2-i\infty}^{-u^2+i\infty} e^{-v} (-v)^{-(N-1)} dv. \end{aligned}$$

The integral in the last expression is Hankel's integral [1]; namely,

$$(10) \quad \frac{1}{\Gamma(Z)} = \frac{i}{2\pi} \int_{-a-i\infty}^{-a+i\infty} e^{-t} (-t)^{-Z} dt, \quad R(Z) > 0, \quad a > 0.$$

Therefore
$$f(u^2) = \frac{1}{\Gamma(N-1)} e^{-u^2} (u^2)^{N-2},$$

and $dF(Z^2) = \frac{N}{2\sigma^2} \frac{1}{\Gamma(N-1)} e^{-\frac{NZ^2}{2\sigma^2}} \left(\frac{NZ^2}{2\sigma^2}\right)^{N-2} dZ^2$; from which

$$(11) \quad dF(Z) = \frac{2 \left(\frac{N}{2\sigma^2}\right)^{N-1}}{\Gamma(N-1)} e^{-\frac{NZ^2}{2\sigma^2}} Z^{2N-3} dZ.$$

(Note that $f(Z)$ is continuous over $0 \leq Z \leq \infty$.)

This expected result has been obtained by Reno and Mowshowitz [2] who employed an extension of the famous Helmert distribution.

Actually, the result is an obvious one and may be argued as follows: Ns_1^2/σ^2 is distributed as χ^2 with $N-1$ degrees of freedom and Ns_2^2/σ^2 is also distributed as χ^2 with $N-1$ degrees of freedom. Hence, the statistic $\frac{N}{\sigma^2} (s_1^2 + s_2^2)$ is, from the additive property of χ^2 , distributed like χ^2 with $2N-2$ degrees of freedom.

We now turn to the general

Case II: $\sigma_1^2 \neq \sigma_2^2$

No generality will be lost by taking $\sigma_1^2 < \sigma_2^2$. In fact, the present attack will hold with obvious modifications provided $\sigma_1^2 < 2\sigma_2^2$.

Recall that

$$(12) \quad f(Z^2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left\{1 - \frac{2\sigma_1^2 it}{N}\right\}^{-1(N-1)} \left\{1 - \frac{2\sigma_2^2 it}{N}\right\}^{-1(N-1)} e^{-iZ^2 t} dt,$$

at all continuity points of $f(Z^2)$.

In a manner analogous to that employed by Hsu [3], we replace

$$\left(1 - \frac{2\sigma_2^2 it}{N}\right) \text{ by } \frac{\sigma_2^2}{\sigma_1^2} \left(1 - \frac{2\sigma_1^2 it}{N}\right) \left\{1 - \frac{1 - \sigma_1^2/\sigma_2^2}{1 - 2\sigma_1^2 it/N}\right\}.$$

Further, since

$$\left| \frac{1 - \sigma_1^2/\sigma_2^2}{1 - 2\sigma_1^2 it/N} \right| < 1,$$

we may write

$$\begin{aligned} \left\{1 - \frac{2\sigma_2^2 it}{N}\right\}^{-1(N-1)} &= \left(\frac{\sigma_1^2}{\sigma_2^2}\right)^{1(N-1)} \left(1 - \frac{2\sigma_1^2 it}{N}\right)^{-1(N-1)} \sum_{r=0}^{\infty} \frac{\Gamma\left(\frac{N-1}{2} + r\right)}{\Gamma\left(\frac{N-1}{2}\right) \Gamma(r+1)} \\ &\quad \cdot \frac{\left(1 - \frac{\sigma_1^2}{\sigma_2^2}\right)^r}{\left(1 - \frac{2\sigma_1^2 it}{N}\right)^r}. \end{aligned}$$

Thus,

$$(13) \quad f(Z^2) = \frac{\left(\frac{\sigma_1^2}{\sigma_2^2}\right)^{\frac{1}{2}(N-1)}}{2\pi} \int_{-\infty}^{\infty} \sum_{r=0}^{\infty} \frac{\left(1 - \frac{\sigma_1^2}{\sigma_2^2}\right)^r}{r\beta\left(\frac{N-1}{2}, r\right)} \left\{1 - \frac{2\sigma_1^2 it}{N}\right\}^{-\frac{1}{2}(N-1-r)} e^{-\frac{1}{2}Z^2} dt,$$

with the understanding that $r\beta\left(\frac{N-1}{2}, r\right) = 1$ for $r = 0$.

We note that the moduli of the terms of the above series are for all t not greater than the corresponding terms of the following convergent series of positive terms:

$$\sum_{r=0}^{\infty} \frac{\left(1 - \frac{\sigma_1^2}{\sigma_2^2}\right)^r}{r\beta\left(\frac{N-1}{2}, r\right)}.$$

Therefore, uniform convergence over $(-\infty, \infty)$ is established. To show that we may integrate over the infinite interval term by term, we observe that

$|S(t) - S_r(t)| \leq \varphi(t)$ for all t and all large r , where

$$S(t) = \left\{1 - \frac{2\sigma_1^2 it}{N}\right\}^{-\frac{1}{2}(N-1)} \left\{1 - \frac{2\sigma_1^2 it}{N}\right\}^{-\frac{1}{2}(N-1)},$$

$S_r(t)$ = the sum of the first $r+1$ terms of the series, and the function $\varphi(t) = \left|1 - \frac{2\sigma_1^2 it}{N}\right|^{-2}$ which is integrable over $(-\infty, \infty)$. That is, $S_r(t)$ converges to $S(t)$ uniformly relative to $\varphi(t)$.² Hence,

$$(14) \quad f(Z^2) = \frac{\left(\frac{\sigma_1^2}{\sigma_2^2}\right)^{\frac{1}{2}(N-1)}}{2\pi} \sum_{r=0}^{\infty} \frac{\left(1 - \frac{\sigma_1^2}{\sigma_2^2}\right)^r}{r\beta\left(\frac{N-1}{2}, r\right)} \int_{-\infty}^{\infty} \left\{1 - \frac{2\sigma_1^2 it}{N}\right\}^{-\frac{1}{2}(N+r-1)} e^{-\frac{1}{2}Z^2} dt.$$

We have already carried out the integration under Case I with the exception that $(N-1)$ should now be replaced by $(N+r-1)$. The distribution of Z^2 will then be given by

$$(15) \quad dF(Z^2) = \left(\frac{\sigma_1^2}{\sigma_2^2}\right)^{\frac{1}{2}(N-1)} \sum_{r=0}^{\infty} \frac{\left(1 - \frac{\sigma_1^2}{\sigma_2^2}\right)^r}{r\beta\left(\frac{N-1}{2}, r\right)} \cdot \frac{\frac{N}{2\sigma_1^2}}{\Gamma(N+r-1)} \cdot e^{-\frac{NZ^2}{2\sigma_1^2}} \left(\frac{NZ^2}{2\sigma_1^2}\right)^{N+r-2} d(Z^2).$$

² The author is indebted to Prof. E. J. McShane for this definition which is due to Prof. E. H. Moore. It may be shown easily that $\lim_{r \rightarrow \infty} \int_{-\infty}^{\infty} S_r(t) dt = \int_{-\infty}^{\infty} \lim_{r \rightarrow \infty} S_r(t) dt$.

Finally, the distribution function of Z is

$$(16) \quad dF(Z) = 2 \left(\frac{\sigma_1^2}{\sigma_2^2} \right)^{1/2(N-1)} \sum_{r=0}^{\infty} \frac{\left(1 - \frac{\sigma_1^2}{\sigma_2^2} \right)^r}{r\beta \left(\frac{N-1}{2}, r \right)} \cdot \frac{\left(\frac{N}{2\sigma_1^2} \right)^{N+r-1}}{\Gamma(N+r-1)} e^{-\frac{1}{2\sigma_1^2} Z^2} Z^{2N+2r-2} dz.$$

We remark that the above series expansion holds, of course, for N odd or even. In case N is odd it may be shown that the distribution function may be expressed as a finite series of Incomplete Gamma Functions.³ However, the finite expansion for N odd appears to offer no marked advantage since for computational purposes the infinite series expansion converges quite rapidly (N either odd or even) and may be put into a convenient form given below.

4. Computational form for the distribution function. In deciding whether or not an observed value of Z is significant and likewise in control chart procedure, one is interested in the percentage points of $f(Z)$. For example, it may be desired to find the value of k such that $P\{Z \leq k\sqrt{\sigma_1^2 + \sigma_2^2}\} = .995$, say, for various sample sizes N . In this connection it will be convenient to work with the distribution of Z^2 , for $P\{Z \leq k\sqrt{\sigma_1^2 + \sigma_2^2}\} = P\{Z^2 \leq k^2(\sigma_1^2 + \sigma_2^2)\}$ also. Now,

$$(18) \quad \begin{aligned} P\{Z^2 \leq k^2(\sigma_1^2 + \sigma_2^2)\} &= \int_0^{k^2(\sigma_1^2 + \sigma_2^2)} dF(Z^2) \\ &= \left(\frac{\sigma_1^2}{\sigma_2^2} \right)^{1/2(N-1)} \sum_{r=0}^{\infty} \frac{\left(1 - \frac{\sigma_1^2}{\sigma_2^2} \right)^r}{r\beta \left(\frac{N-1}{2}, r \right)} \cdot \frac{\left(\frac{N}{2\sigma_1^2} \right)^{N+r-1}}{\Gamma(N+r-1)} \\ &\quad \cdot \int_0^{k^2(\sigma_1^2 + \sigma_2^2)} e^{-\frac{N}{2\sigma_1^2} Z^2} \left(\frac{NZ^2}{2\sigma_1^2} \right)^{N+r-2} d(Z^2), \end{aligned}$$

since we may integrate the series term by term over the entire range of Z^2 or any part of it [5]. In the terminology of Karl Pearson's Incomplete Gamma Function [3],

$$(19) \quad I(u, p) = \frac{1}{\Gamma(p+1)} \int_0^{u\sqrt{v+1}} e^{-v} v^p dv,$$

we may write the above series in the form

$$(20) \quad \begin{aligned} P\{Z^2 \leq k^2(\sigma_1^2 + \sigma_2^2)\} &= \left(\frac{\sigma_1^2}{\sigma_2^2} \right)^{1/2(N-1)} \sum_{r=0}^{\infty} \frac{\left(1 - \frac{\sigma_1^2}{\sigma_2^2} \right)^r}{r\beta \left(\frac{N-1}{2}, r \right)} I \left\{ \frac{Nk^2 \left(1 + \frac{\sigma_2^2}{\sigma_1^2} \right)}{2\sqrt{N+r-1}}, N+r-2 \right\}. \end{aligned}$$

³ Prof. C. C. Craig kindly pointed out this fact to the author.

It is indeed convenient and enlightening that the result is a function of the ratio, σ_1^2/σ_2^2 , and not σ_1^2 and/or σ_2^2 explicitly.

Hence, for a given sample size and ratio of σ_1^2/σ_2^2 , we may find k by inverse interpolation such that $P\{Z \leq k\sqrt{\sigma_1^2 + \sigma_2^2}\} = \alpha$, any desired level of probability.

5. Moments and percentage points for Case I. For the case met many times in practice, i.e. $\sigma_1^2 = \sigma_2^2 = \sigma^2$, we will give a table of the mean and standard deviation and also several probability levels which are obtainable directly from the percentage points of the χ^2 distribution [6].

From (11), we have

$$(21) \quad E(Z^k) = \frac{2 \left(\frac{N}{2\sigma^2} \right)^{N-1}}{\Gamma(N-1)} \int_0^\infty e^{-\frac{NZ^2}{2\sigma^2}} Z^{2N+k-1} dZ \\ = \frac{\Gamma(N-1+k/2)}{\Gamma(N-1)} \left(\frac{2\sigma^2}{N} \right)^{k/2}.$$

Thus,

$$(22) \quad \mu'_{1;Z} = \frac{\Gamma(N-1/2)}{\Gamma(N-1)} \sqrt{\frac{2}{N}} \sigma,$$

$$(23) \quad \mu'_{2;Z} = \frac{2(N-1)}{N} \sigma^2,$$

and

$$(24) \quad \mu_{2;Z} = \frac{2}{N} \left\{ N-1 - \left[\frac{\Gamma(N-1/2)}{\Gamma(N-1)} \right]^2 \right\} \sigma^2.$$

In the table below, the mean and standard deviation are given as a multiple of $\sqrt{2}\sigma$ and $k_{.95}$, for example, is that value of k such that $P\{Z \leq k\sqrt{2}\sigma\} = .95$.

TABLE I

N	Mean	Standard Deviation	Percentage Points			
			$k_{.005}$	$k_{.01}$	$k_{.05}$	$k_{.995}$
2	.6267	.3276	.0501	.1602	1.2239	1.6276
3	.7675	.2780	.1857	.3442	1.2575	1.5738
4	.8308	.2443	.2906	.4521	1.2546	1.5226
5	.8670	.2198	.3667	.5227	1.2453	1.4817
6	.8904	.2014	.4239	.5730	1.2351	1.4488
7	.9068	.1869	.4680	.6110	1.2255	1.4218
8	.9189	.1752	.5046	.6408	1.2167	1.3991
9	.9282	.1653	.5345	.6651	1.2087	1.3798
10	.9356	.1569	.5597	.6852	1.2014	1.3630
11	.9416	.1498	.5813	.7023	1.1949	1.3483
12	.9466	.1434	.6001	.7170	1.1889	1.3353
13	.9508	.1378	.6166	.7298	1.1835	1.3237
14	.9544	.1330	.6313	.7411	1.1784	1.3132
15	.9575	.1285	.6445	.7512	1.1738	1.3038

REFERENCES

- [1] WHITTAKER AND WATSON, *Modern Analysis*, Fourth Edition, 1922, pp. 244-246.
- [2] F. V. RENE AND SIMON MOWSHOWITZ, "The distribution of the radial standard deviation", Ballistic Research Laboratory Report No. 322, Aberdeen Proving Ground, Md.
- [3] P. L. HAY, "Contribution to the theory of student's *t*-Test as applied to the problem of two samples", *Statistical Research Memoirs*, Vol. 2, December 1938, pp. 1-24.
- [4] KARL PEARSON (Editor), *Tables of the Incomplete Gamma Function*, Cambridge University Press, 1934.
- [5] BROMWICH, *An Introduction to the Theory of Infinite Series*, MacMillan, 1908, pp. 453-454.
- [6] CATHERINE M. THOMPSON, "Table of percentage points of the χ^2 distribution", *Biometrika*, Vol. 32, Part II (1941), pp. 187-191.

A MATRIX PRESENTATION OF LEAST SQUARES AND CORRELATION THEORY WITH MATRIX JUSTIFICATION OF IMPROVED METHODS OF SOLUTION

BY PAUL S. DWYER

University of Michigan

1. Introduction and summary. It is the aim of this paper to exhibit, by using elementary matrix theory, the basic concepts of least squares and correlation theory, the solution of the normal equations, and the presentation and justification of recently developed and newly proposed techniques into a single, compact, and short presentation. We shall be mainly concerned with the following topics:

a. Basic least squares theory including derivation of normal equations, the theoretical solution of these equations (regression coefficients), the standard errors of these solutions, and the standard error of estimate.

b. The more specific theory (correlation theory) resulting from applying the general least squares results to the standardized distributions.

c. A matrix presentation of the Doolittle solution.

d. A simple matrix justification of methods, previously presented, for getting least squares and multiple correlation constants from the entries of an abbreviated Doolittle solution.

e. A presentation of a more general theory which the matrix presentation reveals.

f. The outline of a "square root" method as an alternative to the Doolittle method.

The reader should be familiar with elementary matrix theory such as that outlined on pages 1-57 of Aitken's book [1].

No previous knowledge of the Doolittle technique is demanded although a familiarity with the notation and contents of two earlier papers [2], [3] is advised, particularly for those who are interested in the computational aspects.

The presentation here is theoretical and is not concerned with such computational topics as the number of decimal places required, etc. With reference to the number of places, the reader is referred to the recent paper of Professor Hotelling [4].

2. Notation. Let $[x'_i]$ with $1 \leq i \leq N$ and $1 \leq j \leq n$ be the n by N matrix of observed variates of n "predicting variables" for N individuals with i indicating the individual and j the variable. Let $[y'_i]$ be the one by N column matrix of the observed variates of the "predicted" variable. Let the matrices of deviations from the variable means be indicated by $[x_i] = X$ and $[y_i] = Y$. Then by the least squares hypothesis we are to find numbers $b_{y1}, \dots, b_{y2}, \dots, \dots, b_{yn}, \dots$ such that

$$e_i = y_i - (x_{i1}b_{y1} + x_{i2}b_{y2} + \dots + x_{in}b_{yn}),$$

shall have a minimum variation (standard deviation). We then denote the b_{vi} by the one by n column matrix B and the e_i by the one by N column matrix E and have

$$(1) \quad E = Y - XB,$$

as the basic matrix equation.

It may be noted further that the fitted values of y_i are given by the one by N matrix product $XB = Y$. Using this notation (1) appears as

$$(1') \quad E = Y - Y.$$

3. Basic least squares theory.

a. *Sum of squares of residuals.* The condition for minimum variation in this situation (variates measured from means) is equivalent to the condition for minimum sum of squares of residuals. In matrix notation this sum of the squares of the residuals can be written

$$(2) \quad E'E \text{ with } E = Y - XB = Y - Y, \text{ and } E' \text{ the transpose of } E.$$

b. *The normal equations.* Differentiating (2) with respect to B' we find the necessary condition to be

$$(3) \quad X'E = 0.$$

This matrix equation gives the normal equations in implicit form. More explicitly by (1) we have $X'(Y - XB) = 0$ so

$$(4) \quad X'XB = X'Y.$$

The reader should immediately recognize that (4) is the matrix equivalent of the usual statement of the normal equations where deviations from the means are used. It should be noted also that (3) and (4) can be written in the form

$$(5) \quad X'Y = X'Y \text{ from whence at once } Y'Y = Y'Y.$$

c. *Solution of normal equations.* The theoretical solution of (4) is accomplished at once and results in

$$(6) \quad B = (X'X)^{-1}X'Y = (X'X)^{-1}X'Y.$$

d. *Standard deviation of residuals.* The standard deviation of residuals is $\sqrt{E'E/N}$.

In order to evaluate this we note that

$$(7) \quad Y'E = B'X'E = 0, \quad \text{and } Y'Y = Y'Y,$$

Thus

$$(8) \quad E'E = (Y - Y)'E = Y'E = Y'(Y - XB) = Y'Y - Y'XB,$$

and

$$(9) \quad Y'XB = Y'Y = Y'Y = Y'Y.$$

Since $Y'Y = \Sigma y^2$, we have

$$(10) \quad E'E = \Sigma y^2 \left(1 - \frac{Y'XB}{\Sigma y^2} \right),$$

so that, dividing by N and taking the square root

$$(11) \quad s_e = s_y \sqrt{1 - \frac{Y'XB}{\Sigma y^2}}.$$

If the relation between the estimated standard deviations in the population is desired then divide each side of (10) by the number of degrees of freedom and get

$$(12) \quad \sigma_e = \sigma_y \sqrt{1 - \frac{Y'XB}{\Sigma y^2}}.$$

Alternative formulas to (11) and (12) are obtained by replacing $Y'XB$ by its equivalent expressions in (9).

e. Formulas for multiple correlation coefficient. It is to be noted that the numerical quantity $Y'XB/\Sigma y^2$ plays an important role in measuring the ratio σ_e/σ_y . It is customary to use this quantity as the definition of the square of the multiple correlation coefficient so we have

$$(13) \quad r_{y \cdot x_1 \cdot x_n}^2 = \frac{Y'XB}{\Sigma y^2} = \frac{B'X'XB}{\Sigma y^2} = \frac{Y'X(X'X)^{-1}X'Y}{\Sigma y^2} = \frac{Y'Y}{\Sigma y^2} = \frac{\sigma_y^2}{\sigma_e^2}.$$

f. Formulas for correlation coefficient. When $n = 1$, $X'X = \Sigma x^2$, $Y'X = X'Y = \Sigma XY$, $B = b$ and (13) gives

$$(14) \quad r_{yx} = \sqrt{\frac{b\Sigma xy}{\Sigma y^2}} = b \sqrt{\frac{\Sigma x^2}{\Sigma y^2}} \left(= b \frac{\sigma_x}{\sigma_y} \right) = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{\sigma_{xy}}{\sigma_y}.$$

Many of the above developments can be duplicated, without formal use of matrix theory, by judicious use of symbolism and substitution. See for example the presentations of Kirkham [5], Bacon [6], and Guttman [7].

g. Errors of regression coefficients. If B_0 is an approximation to B such that $B_0 + \Delta B = B$ then (6) can be written

$$B_0 + \Delta B = (X'X)^{-1}X'Y$$

and

$$(15) \quad \Delta B = (X'X)^{-1}X'(Y - XB_0).$$

This formula can be used in finding corrections ΔB necessary to change any proposed trial solution, B_0 , into a correct solution. It could also be used in extending the accuracy of a solution after an approximation had been secured to a specific number of places. It has greater utility however in another problem.

We suppose that the predicting variables, the x 's, contain no errors but that there are errors in the observed values of y . Let the hypothetical observed values of y be indicated by Y and the recorded observed values of y by Y_0 . Let the values of B_0 be the regression coefficients obtained by using the recorded observed values Y_0 . Thus $\Delta B = 0$ when Y is replaced by Y_0 in (15). Now let $Y = XB_0$, the "true" residual errors of the recorded observed values, be indicated by E . Then (15) becomes

$$(16) \quad \Delta B = (X'X)^{-1}X'E.$$

Sampling theory can be applied to (16) to obtain a formula for the standard error of the regression coefficient. It is assumed that the "true" residual errors are independent with a common standard deviation σ_e . The values of ΔB are then linear functions of these errors. It follows that

$$(17) \quad \sigma_B^2 = \sigma_{\Delta B}^2 = (X'X)^{-1}X'X(X'X)^{-1}\sigma_e^2 = (X'X)^{-1}\sigma_e^2.$$

The standard errors of the regression coefficients are thus formed by multiplying σ_e by the square roots of the diagonal terms of the inverse of $X'X$.

4. Standard variates. Use of correlation matrix. Many of the formulas of section 3 are simplified with the use of some type of standardization. In particular it is possible to reduce the matrix $X'X$ to the matrix R of correlation coefficients by replacing x by t_x/N where $t_x = x/s$. If y is similarly replaced and B by B , then $X'Y = R_{xy}$ and $Y'X = R'_{xy}$, $Y'Y = \Sigma y^2 = 1$ and selected formulas from section 3 become

$$(18) \quad RB = R_{xy}$$

$$(19) \quad B = R^{-1}R_{xy}$$

$$(20) \quad r_{y,x_1 \dots x_n}^2 = R'_{xy}B = B'RB = R'_{xy}R^{-1}R_{xy}.$$

Classical multiple correlation formulas, determinantal and otherwise, are "covered" by the matrix formulas (20).

5. Matrix presentation of a Doolittle solution. Least squares and correlation constants can also be obtained from the entries of a Doolittle solution. We first outline a matrix description of the Doolittle solution of the equation $AX = G$ with $A = [a_{ij}]$ symmetric and of order n .

Let S_1 be a $(n$ by $n)$ matrix with the first row composed of the elements a_{1j} and all other elements 0. Let T_1 be a similar matrix with first row elements $b_{1j} = a_{1j}/a_{11}$ and all other elements 0. Then $A - S_1'T_1 = A_1 = [a_{ij,1}]$ is a symmetric $(n$ by $n)$ matrix with all elements of the first row and the first column 0.

Next let S_2 be a $(n$ by $n)$ matrix with second row elements $a_{2j,1} = a_{2j} - a_{21}b_{1j}$ and all other elements 0. Let T_2 be a $(n$ by $n)$ matrix with second row elements $b_{2j,1} = a_{2j,1}/a_{22,1}$ and all other elements 0. Then it follows that the matrix

$A_1 - S'_1T_1 = [a_{1,12}]$ is a symmetric (n by n) matrix with the elements of the first 2 columns and the first 2 rows all 0.

This process is continued through successive steps, an additional row and column being made identically 0 at each step, through n steps. At the end of n steps we have the result.

$$(21) \quad A - S'_1T_1 - S'_2T_2 - \cdots - S'_nT_n = 0.$$

This development, when applied to each side of the matrix equation, provides the basis for an equation solving technique which Aitken has called the "method of pivotal condensation" (8) but which the author feels is more adequately characterized as the "method of single division" (9). The Abbreviated Doolittle method can be obtained as an abbreviation of this method. It is not necessary to compute all the elements of the successive matrices $A_1A_2 \cdots$, etc. but only the non-zero elements of the $S_1, T_1, S_2, T_2 \cdots$ etc. matrices.

Consider the so called triangular matrix $S = S_1 + S_2 + S_3 + \cdots + S_n$ with its rows composed of the non-zero rows of the S_j . Consider also the matrix $T = T_1 + T_2 + \cdots + T_n$. Then

$$(22) \quad S'T = S'_1T_1 + S'_2T_2 + \cdots + S'_nT_n$$

since $S'_iT_j = 0$ when $i \neq j$.

It follows that (21) can be written

$$(23) \quad A - S'T = 0.$$

An efficient way of building up these matrices S and T in practice and in making the corresponding transformations on the right side of the equation is the Abbreviated Doolittle method. It is apparent from (23) that the Doolittle method is directed, in part at least, toward the factorization of the symmetric matrix A into two triangular matrices.

It should be noted that these triangular matrices are related by the matrix formula

$$(24) \quad S = DT,$$

where D is the diagonal matrix with diagonal elements $a_{11}, a_{22}, a_{33}, \cdots, a_{nn}$.

Operations performed on the left of the matrix equations $AX = G$ are also performed on the right side so that the Doolittle technique results in the establishment of the auxiliary matrix equations.

$$(25) \quad SX = SA^{-1}G.$$

$$(26) \quad TX = TA^{-1}G.$$

A simple outline ($n = 3$) of the form of the Abbreviated Doolittle method is presented for the purpose of identifying these matrices. A is symmetric and G is the column matrix $[a_{14}]$.

a_{11}	a_{12}	a_{13}	a_{14}
	a_{22}	a_{23}	a_{24}
		a_{33}	a_{34}
a_{11}	a_{12}	a_{13}	a_{14}
1	b_{12}	b_{13}	b_{14}
	$a_{22 \cdot 1}$	$a_{23 \cdot 1}$	$a_{24 \cdot 1}$
	1	$b_{23 \cdot 1}$	$b_{24 \cdot 1}$
		$a_{33 \cdot 12}$	$a_{34 \cdot 12}$
		1	$b_{34 \cdot 12}$

The matrix S is then $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22 \cdot 1} & a_{23 \cdot 1} \\ 0 & 0 & a_{33 \cdot 12} \end{bmatrix}$, the matrix T is

$$\begin{bmatrix} 1 & b_{12} & b_{13} \\ 0 & 1 & b_{23 \cdot 1} \\ 0 & 0 & 1 \end{bmatrix}, \quad SA^{-1}G \text{ is } \begin{bmatrix} a_{14} \\ a_{24 \cdot 1} \\ a_{34 \cdot 12} \end{bmatrix}, \quad TA^{-1}G \text{ is } \begin{bmatrix} b_{14} \\ b_{24 \cdot 1} \\ b_{34 \cdot 12} \end{bmatrix}.$$

6. Least squares and multiple correlation constants from the Doolittle solution. The inverse of A is needed for many formulas. We set up a technique for solving $AY = I$ simultaneously with $AX = G$. This is indicated symbolically by

$$\begin{array}{c|c|c} A & G & I \\ \hline S & SA^{-1}G & SA^{-1} \\ \hline T & TA^{-1}G & TA^{-1} \end{array}.$$

It follows at once that

$$(27) \quad (SA^{-1})'(TA^{-1}) = A^{-1}S'TA^{-1} = A^{-1}AA^{-1} = A^{-1}.$$

This matrix multiplication is easily and readily accomplished when the matrices are in the Doolittle form.

Similarly

$$(28) \quad (SA^{-1})'(TA^{-1}G) = A^{-1}G = X, \text{ the matrix of solutions of } AX = G \text{ and}$$

$$(29) \quad (SA^{-1}G)'(TA^{-1}G) = G'A^{-1}G.$$

It is interesting to note further that $(SA^{-1})'$ and T are inverse triangular matrices since

$$(30) \quad (SA^{-1})'T = A^{-1}S'T = I.$$

S' and TA^{-1} have a similar relationship.

In the case of least squares theory $A = X'X$, $G = X'Y$, $X = B$ so that the formulas (27)(28)(29) become

$$(31) \quad (SA^{-1})'(TA^{-1}) = A^{-1} = (X'X)^{-1}.$$

$$(32) \quad (SA^{-1})'(TA^{-1}G) = X = B.$$

$$(33) \quad (SA^{-1}G)'(TA^{-1}G) = G'A^{-1}G = Y'X(X'X)^{-1}X'Y \\ = Y'XB = B'X'XB = Y'Y.$$

If the normal equations are reduced to standard form $A = R$, $X = B$, $G = R_{xy}$ and we have

$$(34) \quad (SA^{-1})'(TA^{-1}) = A^{-1} = R^{-1}.$$

$$(35) \quad (SA^{-1})'(TA^{-1}G) = X = B.$$

$$(36) \quad (SA^{-1}G)'(TA^{-1}G) = G'A^{-1}G = R'_{xy}R^{-1}R_{xy} = R'_{xy}B = B'RB = r_{yx}^2 \dots r_{xn}^2.$$

The reader is referred to an earlier paper [3, 457] for an illustration of these techniques.

It should be noted that the solution is a cumulative one in the sense that solutions involving n predicting variables are obtained from solutions involving $n - 1$ predicting variables by the addition of paired products. This is a highly desirable feature as it makes possible direct analyses showing the effect of an added predicting variable.

7. A more general theory—solution of matrix equations by factorization. Examination of the results of section 5 leads one at once to a consideration of a more general theory. The key formula in this development is $A - S'T' = 0$ and all subsequent formulas stem from this. Hence if A can be factored into any matrices, S' and T , not necessarily triangular, the results of section 6 follow.

From a practical standpoint it is desired that the factorization process yield, simultaneously, the values S , T , $SA^{-1}G$; $TA^{-1}G$; SA^{-1} and TA^{-1} as the Doolittle method does. But, formally, these can be computed if S and T are known.

8. A "square root" method. A most interesting and practical special case of the above method is that in which the triangular matrices S and T are equal. It appears that a technique based on this property would have some advantages over the Doolittle method since the double rows of the Doolittle solution could be replaced by single rows, while the formulas of sections 5 and 6 are just as applicable. Now such a technique is easily devised. From (23) and (24) we see that

$$(37) \quad A - S'D^{-1}S = 0,$$

where D is a diagonal matrix.

We replace $D^{-1}S$ by a new S , $(D^{-1}S)'$ by a new S' and have

$$(38) \quad A - S'S = 0.$$

The technique of solution is similar to that of the Doolittle except that the entries s_{ij} are $a_{ij} / \sqrt{a_{ii}}$. These values s_{ij} are thus geometric means of the values a_{ij} and b_{ij} .

A simple machine technique is available for computing these entries. In some respects the solution is superior to the Doolittle solution. It is hardly pertinent to the subject matter of this paper to present a detailed discussion of the merits of this method, with the numerical illustrations. This will be done in a later paper.

After arriving at this method by the steps described above, it seemed surprising that such a simple and compact method has not been discovered by some previous worker. Although matrix factorization is not a new subject, I have not found evidence that it has been utilized so directly in the problem of solving matrix equations. The nearest approach I have discovered is the paper by Banachiewicz [10], in which a "square root" method is used in factoring A .

REFERENCES

- [1] A. C. AITKEN, *Determinants and Matrices*, Oliver and Boyd, London, 1942
- [2] P. S. DWYER, "The Doolittle technique," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 449-458.
- [3] P. S. DWYER, "Recent developments in correlation technique," *Jour. Amer. Stat. Assn.*, Vol. 37 (1942), pp. 441-460.
- [4] HAROLD HOTELLING, "Some new methods in matrix calculation," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 1-34.
- [5] W. KIRKHAM, "Note on the derivation of the multiple correlation coefficient," *Annals of Math. Stat.*, Vol. 8 (1937), pp. 68-71.
- [6] H. M. BACON, "Note on a formula for the multiple correlation coefficient," *Annals of Math. Stat.*, Vol. 9, (1938), pp. 227-229.
- [7] LOUIS GUTTMAN, "A note on the derivation of a formula for multiple and partial correlation," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 305-308.
- [8] A. C. AITKEN "Studies in practical mathematics I The evaluation, with application, of a certain triple product matrix," *Proc. Royal Soc., Edinburgh*, Vol. 57 (1937), pp. 172-181.
- [9] P. S. DWYER, "The solution of simultaneous equations," *Psychometrika*, Vol. 6 (1941), pp. 101-129.
- [10] T. BANACHIEWICZ, "An outline of the Cracovian algorithm of the method of least squares," *Astr. Jour.*, Vol., 50 (1942), pp. 38-41.

ON THE STATISTICS OF SENSITIVITY DATA

BY BENJAMIN EPSTEIN AND C. WEST CHURCHMAN

Frankford Arsenal

1. **Introduction.** "Sensitivity data" is a general term for that type of experimental data for which the measurement at any point in the scale destroys the sample; as a consequence, new samples are required for each determination. Examples of such data occur in biology in dosage-mortality determinations, in psychophysics in questions concerning sensitivity responses, and, more recently, in the theory of solid explosives, in questions concerning the sensitivity of explosive or detonative mixtures.

Methods of analyzing such data have been discussed by Bliss¹ and Spearman², and others. The present paper is a generalization of Spearman's result; it is the feeling of the authors that Spearman's method, if properly founded in mathematical theory, is preferable to Bliss', for it does not necessitate the assumption of some type of distribution prior to analysis, and hence resembles the standard treatment of independent observations made on the same object.

Throughout the following discussion, we let x_i be the magnitude of a certain "stimulus" (be it dosage, physical stimulus, or strength of blow) and p_i the corresponding fraction of objects unaffected by the stimulus. Bliss' method consisted in assuming that the p_i represented the cumulative distribution of some known function (in his case, the normal function), and hence the p_i could be transformed into a variable t_i linearly dependent on the x_i . The difficulty of this treatment, in addition to the distribution assumption, lies in the fact that the t_i do not have equal standard errors, and the straight line fit is very cumbersome.

Instead, Spearman makes the much simpler assumption that if p_i is unaffected at x_i , and p_{i+1} at x_{i+1} , then $p_i - p_{i+1}$ is an estimate of the fraction that is just affected (i.e., the fraction of those that have "critical" responses) at about $\frac{1}{2}(x_i + x_{i+1})$. If the x_i are evenly spaced, as we shall assume them to be throughout, and $p_1 = 1.0$ and $p_n = 0$, then any set of sensitivity data may be transformed into a set of data on critical responses classified into classes whose mid-points are evenly spaced. Without loss of generality, we shall assume the x_i 's to be integers and the intervals to be unity. The data on critical responses can then be treated in the normal way, and \bar{X} and all the measures of dispersion calculated in the usual fashion. In order to justify such procedures, however, it is necessary to show how the sampling errors of \bar{X} and the higher moments can be estimated.

¹ C. I. Bliss, "The calculation of the dosage mortality curve," *Annals of Applied Biology*, Vol. 22, pp. 134-187.

² C. Spearman, "The method of 'right and wrong cases' (constant stimuli) without Gauss' formulae," *British Jour. of Psych.*, Vol. 2, 1908, pp. 227-242.

2. The moments and their errors. By definition,

$$(1) \quad \bar{X} = \sum_{i=1}^n (p_i - p_{i+1})(x_i + x_{i+1})/2 = \sum_{i=1}^n (p_i - p_{i+1})(x_i + \frac{1}{2}).$$

If we let x_1 represent the stimulus for which none of the samples can be affected, then

$$(2) \quad \bar{X} = x_1 + .5 + \sum_{i=2}^{n-1} p_i,$$

as Spearman has shown (3). Since x_1 is constant, and the p_i are all independent (non-correlated), it follows that (N_i being the number of objects in the i th sample)

$$(3) \quad \sigma_x^2 = \sigma_{\Sigma p_i}^2 = \sigma_{p_1}^2 + \sigma_{p_2}^2 + \dots + \sigma_{p_n}^2 = \sum \sigma_{p_i}^2 = \sum_{i=2}^{n-1} \frac{p_i q_i}{N_i}$$

(since $\sigma_{p_1}^2 = \sigma_{p_n}^2 = 0$).

Again by definition, the q th moment about the origin is

$$(4) \quad \mu'_q = \sum_{i=1}^n (p_i - p_{i+1})(x_i + \frac{1}{2})^q.$$

As before $x_1 + .5$ can be taken as the origin ($x_1 + .5 = 0$), in which case we have

$$(5) \quad \mu'_q = (p_1 - p_2) \cdot 0^q + (p_2 - p_3) \cdot 1^q + (p_3 - p_4) \cdot 2^q \\ + \dots + (p_{n-1} - p_n)(n-1)^q.$$

If we let $b_{q,i}$ represent the i th first difference of the consecutive q th powers of the positive integers (including 0), then

$$(6) \quad \mu'_q = \sum_{i=2}^{n-1} b_{q,i} p_i,$$

by expansion of (5). Hereafter all Σ will be taken from $i = 2$ to $i = n - 1$.

Evidently

$$(7) \quad \sigma_{\mu'_q}^2 = \sum_{i=2}^{n-1} b_{q,i}^2 \left(\frac{p_i q_i}{N_i} \right),$$

or

$$(8) \quad \sigma_{\mu'_q}^2 = \sum_{i=2}^{n-1} b_{q,i}^2 \sigma_{p_i}^2.$$

We are interested now in the standard error of the q th moment about the sample mean. To obtain this, compute first the correlation between the q th and r th moments about the origin.

If $\delta\mu'_q$ is taken to be the sample error in μ'_q due to deviations δp_i from the true values, then we have

$$\delta\mu'_q = \sum_{i=2}^{q-1} b_{q,i} \delta p_i,$$

$$\delta\mu'_r = \sum_{i=2}^{r-1} b_{r,i} \delta p_i.$$

Hence

$$\delta\mu'_q \delta\mu'_r = \sum b_{q,i} b_{r,i} (\delta p_i)^2 + \sum_{i,j} (b_{q,i} b_{r,j} + b_{q,j} b_{r,i}) \delta p_i \delta p_j.$$

Summing for all samples:

$$\begin{aligned} \sigma_{\mu'_q} \sigma_{\mu'_r} r_{\mu'_q \mu'_r} &= \sum b_{q,i} b_{r,i} \sigma_{p_i}^2 + \sum_{i,j} (b_{q,i} b_{r,j} + b_{q,j} b_{r,i}) (\sigma_{p_i} \sigma_{p_j} r_{p_i p_j}) \\ (9) \qquad \qquad \qquad &= \sum b_{q,i} b_{r,i} \sigma_{p_i}^2. \end{aligned}$$

Since evidently $r_{p_i p_j}$ vanishes for all $i \neq j$ (the p_i being completely independent in the statistical sense),

In particular, when $\mu'_r = \mu'_1 = \bar{X}$, we have

$$(10) \qquad \qquad \qquad \sigma_{\mu'_q} \sigma_{\bar{X}} r_{\mu'_q \bar{X}} = \sum b_{q,1} \sigma_{p_1}^2.$$

By definition, the q th moment about the mean will be

$$(11) \qquad \mu_q = \Sigma (p_i - p_{i+1}) (x_i + \tfrac{1}{2} - \bar{X})^q = \Sigma p'_i (x_i + \tfrac{1}{2} - \bar{X})^q$$

where $p'_i = p_i - p_{i+1}$.

For computational purposes, this may be written as

$$(12) \quad \mu_q = \mu'_q - q \bar{X} \mu'_{q-1} + \frac{q(q-1)}{2} \bar{X}^2 \mu'_{q-2} + \dots + {}_q C_r \bar{X}^r \mu'_{q-r+1} + \dots + \bar{X}^q$$

where $\bar{X} = \Sigma p_i = \mu'_1$, if $x_1 + \frac{1}{2}$ is the origin.

To obtain $\sigma_{\mu'_q}^2$, where \bar{X} is estimated from the sample, we may follow the usual procedures, arguing that

$$(13) \qquad \delta\mu_q = \Sigma \{ (x_i + \tfrac{1}{2})^q \delta p'_i \} - q \delta \bar{X} \Sigma (x_i + \tfrac{1}{2})^{q-1} p'_i + T$$

where T contains terms involving \bar{X} and higher powers of \bar{X} .

From (13) we obtain

$$(14) \qquad \sigma_{\mu_q}^2 = \sigma_{\mu'_q}^2 + q^2 \mu'_{q-2} \sigma_{\bar{X}}^2 - 2q \mu'_{q-1} \sigma_{\bar{X}} r_{\mu'_q \bar{X}} + U$$

where U involves \bar{X} and higher powers. From (3), (8), (10) and (14) we have

$$\begin{aligned} (15) \qquad \sigma_{\mu_q}^2 &= \Sigma b_{q,i}^2 \sigma_{p_i}^2 + q^2 \mu'_{q-2} \sigma_{p_1}^2 - 2q \mu'_{q-1} \Sigma b_{q,i} \sigma_{p_i}^2 + U \\ &= \Sigma (b_{q,i} - q \mu'_{q-1})^2 \sigma_{p_i}^2 + U. \end{aligned}$$

We now shift the origin to \bar{X} . All the terms in U vanish, μ'_{q-1} becomes μ_{q-1} , and the $b_{q,i}$ values go into $\beta_{q,i}$, where

$$\beta_{q,i} = (i - \bar{X})^q - (i - 1 - \bar{X})^q.$$

That is, (15) becomes

$$(16) \quad \sigma_{\mu_q}^2 = \Sigma(\beta_{q,i} - q\mu_{q-1})^2 \sigma_{p,i}^2.$$

It is of interest to give an alternative proof of the relation (16) possessing the desirable property of being very short and simple and at the same time yielding an expression for the $\beta_{q,i}$ in terms of $b_{r,i}$ ($1 \leq r \leq q$) and powers of \bar{X} .

If $x_1 + 5$ is taken as the origin, then (11) may be written as (17)

$$(17) \quad \mu_q = \Sigma(i - \bar{X})^q p'_i.$$

The application of the δ -operation to both sides of (17) yields:

$$(18) \quad \delta\mu_q = \Sigma(i - \bar{X})^q \delta p'_i - q \Sigma(i - \bar{X})^{q-1} p'_i \delta \bar{X}$$

$$(19) \quad = \Sigma(\beta_{q,i} - q\mu_{q-1}) \delta p_i.$$

Repetition of a previous argument gives the result:

$$(20) \quad \sigma_{\mu_q}^2 = \Sigma(\beta_{q,i} - q\mu_{q-1})^2 \sigma_{p,i}^2.$$

In order to derive the relation connecting the $\beta_{q,i}$ with $b_{r,i}$ ($1 \leq r \leq q$) we expand $\Sigma(i - \bar{X})^q \delta p'_i$ in equation (18). This expansion yields:

$$(21) \quad \begin{aligned} \Sigma(i - \bar{X})^q \delta p'_i &= \Sigma(i^q - {}_q C_1 i^{q-1} \bar{X} + {}_q C_2 i^{q-2} \bar{X}^2 \\ &\quad + \dots + (-1)^{q-1} {}_q C_{q-1} i \bar{X}^{q-1} + (-1)^q \bar{X}^q) \delta p'_i \\ &= \Sigma(b_{q,i} - {}_q C_1 b_{q-1,i} \bar{X} + {}_q C_2 b_{q-2,i} \bar{X}^2 \\ &\quad + \dots + (-1)^{q-1} q \bar{X}^{q-1} b_{1,i}) \delta p_i \end{aligned}$$

i.e.,

$$(22) \quad \begin{aligned} \beta_{q,i} &= b_{q,i} - {}_q C_1 b_{q-1,i} \bar{X} + {}_q C_2 b_{q-2,i} \bar{X}^2 \\ &\quad + \dots + (-1)^{q-1} q \bar{X}^{q-1} b_{1,i}. \end{aligned}$$

The relationship (16) combined with (22) enables one to compute the standard errors of a number of useful statistics. In particular in case $q = 2$ it follows that

$$(23) \quad \sigma_{\mu_2}^2 = \sigma_{\sigma^2}^2 = \Sigma(b_{2,i} - 2\bar{X})^2 \sigma_{p,i}^2.$$

Combining (23) with the well-known result that

$$(24) \quad \sigma_{\sigma} = \sigma_{\mu_2}/2\sigma$$

we see that

$$(25) \quad \sigma_{\sigma} = \frac{\sqrt{\Sigma \{(2i - 3) - 2\bar{X}\}^2 \sigma_{p,i}^2}}{2\sqrt{\Sigma (2i - 3)p_i - (\Sigma p_i)^2}}.$$

Formula (25) is useful in significance tests involving the standard deviations of sensitivity data.

3. Standard errors of the moments in standard units. We now turn our attention to the derivation of the standard error of the higher moments when

expressed in standard units. Before proceeding with the derivation it is convenient to find the correlation between the q th and r th moments about the mean. This result is an immediate consequence of (19), for since

$$\delta\mu_q = \Sigma\{\beta_{q,i} - q\mu_{q-1}\}\delta p_i$$

and

$$(26) \quad \delta\mu_r = \Sigma\{\beta_{r,i} - r\mu_{r-1}\}\delta p_i$$

it follows that

$$(27) \quad \delta\mu_q\delta\mu_r = \Sigma\{\beta_{q,i} - q\mu_{q-1}\}\{\beta_{r,i} - r\mu_{r-1}\}(\delta p_i)^2 + Z$$

where Z contains terms $\delta p_i\delta p_j (i \neq j)$. Hence, as before,

$$(28) \quad \sigma_{\mu_q\mu_r} = \Sigma\{\beta_{q,i} - q\mu_{q-1}\}\{\beta_{r,i} - r\mu_{r-1}\}\sigma_{p_i}^2.$$

Let us now derive the standard errors of the moments in standard units, i.e., of

$$(29) \quad \alpha_q = \mu_q/\sigma^q,$$

Now in general,

$$(30) \quad \delta\alpha_q = \frac{\sigma^q\delta\mu_q - q\sigma^{q-1}\mu_q\delta\sigma}{\sigma^{2q}} = \frac{\sigma\delta\mu_q - q\mu_q\delta\sigma}{\sigma^{q+1}}$$

and since

$$(31) \quad \delta\mu_2 = 2\sigma\delta\sigma, \quad \text{or} \quad \delta\sigma = \delta\mu_2/2\sigma$$

we have

$$(32) \quad \delta\alpha_q = \frac{2\sigma^2\delta\mu_q - q\mu_q\delta\mu_2}{2\sigma^{q+2}}$$

and hence

$$(33) \quad (\delta\alpha_q)^2 = \frac{4\sigma^4(\delta\mu_q)^2 + q^2\mu_q^2(\delta\mu_2)^2 - 4q\sigma^2\mu_q\delta\mu_q\delta\mu_2}{4\sigma^{2(q+2)}}$$

$$(34) \quad \sigma_{\alpha_q}^2 = \frac{4\mu_2^2\sigma_{\mu_q}^2 + q^2\mu_q^2\sigma_{\mu_2}^2 - 4q\mu_2\mu_q\sigma_{\mu_q\mu_2}}{4\mu_2^{q+2}}.$$

In this case, it follows that

$$(35) \quad \sigma_{\alpha_q}^2 = \frac{4\mu_2^2 \sum (\beta_{q,i} - q\mu_{q-1})^2 \sigma_{p_i}^2 + q^2\mu_q^2 \sum \beta_{2,i}^2 \sigma_{p_i}^2 - 4q\mu_q\mu_2 \sum (\beta_{q,i} - q\mu_{q-1})\beta_{2,i} \sigma_{p_i}^2}{4\mu_2^{q+2}}$$

or

$$(36) \quad \sigma_{\alpha_q}^2 = \frac{\sum (2\mu_2(\beta_{q,i} - q\mu_{q-1}) - q\mu_q\beta_{2,i})^2 \sigma_{p_i}^2}{4\mu_2^{q+2}}$$

If the q th moment about the mean vanishes, then

$$(37) \quad \sigma_{\alpha_q}^2 = \frac{4\mu_2^2 \sum (\beta_{q,i} - q\mu_{q-1})^2 \sigma_{p_i}^2}{4\mu_2^{q+2}} = \frac{\sigma_{\mu_q}^2}{\mu_2^q}.$$

It is readily seen that the standard errors of the skewness and flatness are special cases of formula (36) when $q = 3$ and $q = 4$ respectively.

4. Some minimization problems. In the analysis of sensitivity data it is most desirable to minimize σ_X^2 or σ_{σ}^2 in order to increase the precision of significance tests involving \bar{X} or σ respectively. Therefore, it is of interest to solve the following problem: Suppose that we have a sample of size N which is to be subdivided into n samples of size N_i to be tested at a number of fixed levels $\{x_i\}$ $i = 1, 2 \dots n$, $\sum_{i=1}^n N_i = N$. Then what choice of values $\{N_i\}$ will minimize

$$\sigma_X^2 = \sum_{i=1}^n \frac{p_i q_i}{N_i}, \quad \text{where} \quad \sum_{i=1}^n N_i = N?$$

In order to solve this problem most quickly we use the method of Lagrange multipliers, i.e., we minimize the expression

$$(38) \quad L_1(N_i, \lambda) = \sum_{i=1}^n \frac{p_i q_i}{N_i} + \lambda \left(\sum_{i=1}^n N_i - N \right).$$

Taking the partial derivatives with respect to N_i we obtain the n equations

$$(39) \quad \frac{p_i q_i}{N_i^2} = \lambda, \quad \text{i.e.,} \quad N_i = \frac{\sqrt{p_i q_i}}{\lambda^{1/2}}, \quad i = 1, 2 \dots, n.$$

Summing over all values of i we obtain

$$(40) \quad N = \sum_{i=1}^n \frac{\sqrt{p_i q_i}}{\lambda^{1/2}}, \quad \text{or} \quad \lambda^{1/2} = \frac{N}{\sum_{i=1}^n \sqrt{p_i q_i}};$$

i.e., the best choice of values for $\{N_i\}$ is given by

$$(41) \quad N_i = \frac{N \sqrt{p_i q_i}}{\sum_{i=1}^n \sqrt{p_i q_i}}.$$

The value of σ_X^2 for this choice of the set $\{N_i\}$ is

$$(42) \quad \frac{\left(\sum_{i=1}^n \sqrt{p_i q_i} \right)^2}{N}.$$

It is obvious that this is actually a minimum. In particular, it is less than the value of σ_X^2 for $N_i = N/n$ (the number of groups is n). This follows from the application of Schwartz' inequality to (42), for

$$(43) \quad \frac{\left(\sum_{i=1}^n \sqrt{p_i q_i} \right)^2}{N} \leq \frac{n \sum_{i=1}^n p_i q_i}{N}$$

which equals the value of σ_x^2 for $N_1 = N_2 = \dots = N_n = N/n$. The equality holds if and only if $p_1 = p_2 = \dots = p_n$.

Suppose next that we wish to minimize

$$(44) \quad \sigma_{s^2}^2 = \sum_{i=1}^n \frac{(b_{2,i} - 2\bar{X})^2 p_i q_i}{N'_i} = \sum_{i=1}^n \frac{\beta_{2,i}^2 p_i q_i}{N'_i},$$

where

$$\beta_{2,i} = b_{2,i} - 2\bar{X}.$$

We proceed as before to minimize the expression

$$(45) \quad L_2(N'_i, \lambda) = \sum_{i=1}^n \frac{\beta_{2,i}^2 p_i q_i}{N'_i} + \lambda \left(\sum_{i=1}^n N'_i - N \right).$$

Taking partial derivatives with respect to N'_i we obtain

$$(46) \quad \frac{\beta_{2,i}^2 p_i q_i}{N'^2_i} = \lambda \quad \text{i.e.} \quad N'_i = \frac{|\beta_{2,i}| \sqrt{p_i q_i}}{\lambda^{1/2}}, \quad i = 1, 2, \dots, n$$

or summing over all values of i we obtain

$$(47) \quad N = \sum_{i=1}^n \frac{|\beta_{2,i}| \sqrt{p_i q_i}}{\lambda^{1/2}} \quad \text{or} \quad \lambda^{1/2} = \sum_{i=1}^n \frac{|\beta_{2,i}|}{N} \sqrt{p_i q_i},$$

i.e., the best choice of values for $\{N'_i\}$ is given by

$$(48) \quad N'_i = \frac{N |\beta_{2,i}| \sqrt{p_i q_i}}{\sum_{i=1}^n |\beta_{2,i}| \sqrt{p_i q_i}}.$$

The minimum value of $\sigma_{s^2}^2$ is given by

$$(49) \quad \frac{\left(\sum_{i=1}^n |\beta_{2,i}| \sqrt{p_i q_i} \right)^2}{N}.$$

In practice we desire a set $\{N_i\}$ which will make $\sigma_{\bar{X}}^2$ and $\sigma_{s^2}^2$ small simultaneously. Unfortunately this is not in general possible. In fact, it may be asserted that the set $\{N_i\}$ minimizing $\sigma_{\bar{X}}^2$ will yield a large value of $\sigma_{s^2}^2$ and similarly the set $\{N'_i\}$ minimizing $\sigma_{s^2}^2$ will yield a large value of $\sigma_{\bar{X}}^2$. The reason for this curious behavior lies in the fact that the only difference between the set $\{N_i\}$ and the set $\{N'_i\}$ is the set of numbers $\{|\beta_{2,i}|\} = \{|(2i-3) - 2\bar{X}|\}$. These numbers, however, change the character of the sets $\{N_i\}$ and $\{N'_i\}$. In particular $\{N'_i\}$ takes on its largest values for both small and large values of i , whereas $\{N_i\}$ takes on small values in these regions; $\{N'_i\}$ takes on small values for those values of i which are the integral values closest to $\bar{X} + 3/2$, whereas $\{N_i\}$ takes on large values for such values of i . It is this curious juxtaposition of $\{N_i\}$ and $\{N'_i\}$ that renders it impossible to choose sets of numbers $\{N_i\}$ minimizing $\sigma_{\bar{X}}^2$ and $\sigma_{s^2}^2$ simultaneously.

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

NOTE ON RUNS OF CONSECUTIVE ELEMENTS

By J. WOLFOWITZ

Columbia University

In my paper [1] I did not derive the asymptotic distribution of $W(R)$, an omission which I wish to correct in this note.

Let the stochastic variable $R = (x_1, \dots, x_n)$ be a permutation of the first n positive integers, where each permutation has the same probability $\frac{1}{n!}$. A subsequence $x_{i+1}, x_{i+2}, \dots, x_{i+l}$, is called a run of consecutive elements of length l if:

a) when l' is any integer such that $1 \leq l' < l$,

$$|x_{i+l'} - x_{i+l'+1}| = 1$$

b) when $i > 0, |x_i - x_{i+1}| > 1$

c) when $i + l < n, |x_{i+l} - x_{i+l+1}| > 1$.

Let $W(R)$ be the total number of runs in R . Then $n - W(R)$ is a stochastic variable which, it will be shown, has in the limit the Poisson distribution with mean value 2. More precisely, if $p(w)$ is the probability that $n - W(R) = w$, then

$$(1) \quad \lim_{n \rightarrow \infty} p(w) = \frac{2^w}{e^2 \cdot w!}.$$

PROOF: Define stochastic variables $y_i (i = 1, 2, \dots, n)$, as follows: $y_i = 1$ if x_i is the first element of a run of length 2, $y_i = 0$ otherwise. It is easy to see that the probability that $x_i (i = 1, 2, \dots, n)$ be the initial element of a run of length greater than two is $O\left(\frac{1}{n^2}\right)$ and hence that the probability of the occurrence of a run of length greater than two is $O\left(\frac{1}{n}\right)$. Hence the limiting distribution of $n - W(R)$ is the same as that of

$$y = \sum_{i=1}^n y_i,$$

provided either exists.

The y_i are dependent stochastic variables and almost all (i.e., all with the exception of a fixed number) have the same marginal distribution. We now wish to consider the expression

$$E(y_{i_1}^{\alpha_1} y_{i_2}^{\alpha_2} \dots y_{i_k}^{\alpha_k})$$

(where the symbol E denotes the expectation) for any set of fixed positive integers $k, \alpha_1, \dots, \alpha_k$, and for all k -tuples i_1, i_2, \dots, i_k , with no two elements

equal. Now

$$E(y_{i_1}^{\alpha_1} y_{i_2}^{\alpha_2} \cdots y_{i_k}^{\alpha_k}) = E(y_{i_1} y_{i_2} \cdots y_{i_k})$$

is the probability that $y_{i_1} = 1, y_{i_2} = 1, \dots, y_{i_k} = 1$, simultaneously. This probability is either zero (for example, when $|i_2 - i_1| = 1, |i_3 - i_2| = 1$, etc. or when $i_1 = n$, etc.) or $\left(\frac{2}{n}\right)^k + O\left(\frac{1}{n^{k+1}}\right)$. Moreover, the ratio of the number of k -tuples i_1, i_2, \dots, i_k for which the probability is zero to the number of k -tuples for which the probability is $\left(\frac{2}{n}\right)^k + O\left(\frac{1}{n^{k+1}}\right)$ is $O\left(\frac{1}{n}\right)$. Let $Z_i (i = 1, \dots, n)$ be independent stochastic variables each with the same distribution such that the probability that $Z_i = 1$ is $2/n$ and the probability that $Z_i = 0$ is $(n-2)/n$. It follows readily that the limit, as $n \rightarrow \infty$, of the j th moment ($j = 1, 2, \dots$, ad inf.) of y about the origin, is the same as the limit of the same moment of Z , where

$$Z = \sum_{i=1}^n Z_i.$$

Since the Z_i are independently distributed, and since each can take only the values 0 and 1, the probability of the value 1 being $2/n$, the j th moment of Z about the origin approaches, as $n \rightarrow \infty$,

$$\mu_j = e^{-2} \sum_{i=1}^{\infty} \frac{i^j 2^i}{i!},$$

which is the j th moment about the origin of the Poisson distribution with mean value 2. By the preceding paragraph, μ_j is also the limit of the j th moment of y about the origin. Now von Mises [2] has proved that if the j th moment ($j = 1, 2, \dots$, ad inf.) of a chance variable X_n , ($n = 1, 2, \dots$, ad inf.), approaches, as $n \rightarrow \infty$, the j th moment of a Poisson distribution, then the distribution of X_n approaches the Poisson distribution with corresponding mean value. From this it follows that y has in the limit the distribution (1). We have already shown that y and $n - W(R)$ have the same limiting distribution, so that the required result follows.

REFERENCES

- [1] J. WOLFOWITZ, *Annals of Math. Stat.*, Vol. 13 (1942), p. 247.
- [2] R. v MISES, *Zeitschrift für die angewandte Math. und Mechanik*, Vol. 1 (1921), p. 298.

NOTE ON CONSISTENCY OF A PROPOSED TEST FOR THE PROBLEM OF TWO SAMPLES

BY ALBERT H. BOWKER

Columbia University

Certain tests for the hypothesis that two samples are from the same population assume nothing about the distribution function except that it is continuous. Since the power functions of these tests have not been obtained, optimum

tests are not known. However, one desirable¹ test property, that of "consistency," has been introduced by Wald and Wolfowitz [1]. A test is called consistent if the probability of rejecting the null hypothesis when it is false (the power of the test) approaches one as the sample number approaches infinity. This is a logical extension of the familiar idea of consistency introduced by Fisher. It will be shown that a test recently proposed by Mathisen [2] is not consistent with respect to certain alternatives.

The test proposed by Mathisen [2] may be described briefly as follows: Given two samples, observe the number (m) of elements of the second sample whose values are less than the median of the first sample. The distribution of m is independent of the population distribution under the null hypothesis. Let $P\{m < a\}$ denote the probability of the relation in braces under the null hypothesis. If m_1 and m_2 are significance points ($m_1 > m_2$) such that

$$\begin{aligned} P\{m > m_1\} &= \beta_1 \\ P\{m < m_2\} &= \beta_2 \\ \beta_1 + \beta_2 &= \beta < 1, \end{aligned} \tag{1}$$

the statistic m can be used to test the hypothesis at the significance level β . This is called the case of two intervals. The method is extended by using the two quartiles and the median of the first sample to define four intervals into which the elements of the second sample may fall. If the second sample is of size $4n$ and the number which actually falls in each interval is n_1, n_2, n_3 , and n_4 respectively, the distribution of

$$C = \frac{\sum_{i=1}^4 (n_i - n)^2}{9n^2} \tag{2}$$

is also independent of the population distribution under the null hypothesis. Then if C^* is a significance point, such that

$$P\{C > C^*\} = \beta' < 1, \tag{3}$$

C can be used as a test of the hypothesis at the level β' .

To show that Mathisen's test is not consistent, we shall consider first the case of two intervals. Let X and Y be two independent stochastic variables whose cumulative distribution functions $F(x)$ and $G(x)$ are continuous. Let $x_1 < x_2 \cdots < x_{2n+1}$ and $y_1 < y_2 \cdots < y_{2n}$ be sets of ordered independent observations on X and Y . Then m is such that

$$y_m < x_{n+1} < y_{m+1}.$$

Let m_1 and m_2 be the significance points of the distribution of m , defined by (1). Clearly m_1 and m_2 depend on n . We shall prove that the sequence

$$\frac{m_1(n)}{2n} \tag{4} \quad n = 1, 2, \dots$$

¹For large samples.

converges to $\frac{1}{2}$. Since (4) is bounded, it has at least one limit point. Let h be such a limit point. If $h < \frac{1}{2}$ and $\frac{1}{2} - h = 3\delta$, then there exists a monotonically increasing subsequence of the integers n_1, n_2, \dots and a number N such that for $n_i > N$

$$(5) \quad \left| \frac{m_1(n_i)}{2n_i} - h \right| < \delta.$$

Clearly $m/2n$ converges stochastically to $\frac{1}{2}$. Hence if $0 < \epsilon < 1$ is any arbitrarily small number, we can select n so large that the probability is at least $1 - \epsilon$ that

$$(6) \quad \left| \frac{m}{2n} - \frac{1}{2} \right| < \delta.$$

Hence for n sufficiently large, $P\{m > m_1\}$ is at least $1 - \epsilon$, a contradiction with (1). A similar contradiction appears if $h > \frac{1}{2}$. Hence (4) has only one limit point, $\frac{1}{2}$. In the same way we can prove that the sequence

$$(7) \quad \frac{m_2(n)}{2n} \quad n = 1, 2, \dots$$

also converges to $\frac{1}{2}$.

Let $0 < \delta \leq \frac{1}{6}$. Consider now two pairs of populations, A and B, described as follows:

A)	$F(x) \equiv G(x) \equiv x$	$(0 \leq x \leq 1)$
B)	$F(x) \equiv x$	$(0 \leq x \leq 1)$
	$G(x) \equiv 0$	$(0 \leq x \leq \frac{1}{2} - 2\delta)$
	$G(x) \equiv (x - \frac{1}{2} + 2\delta)(\frac{1}{2} - \delta)/\delta$	$(\frac{1}{2} - 2\delta \leq x \leq \frac{1}{2} - \delta)$
	$G(x) \equiv x$	$(\frac{1}{2} - \delta \leq x \leq \frac{1}{2} + \delta)$
	$G(x) \equiv \frac{1}{2} + \delta$	$(\frac{1}{2} + \delta \leq x \leq 1 - \delta)$
	$G(x) \equiv (\frac{1}{2} + \delta) + (x - 1 + \delta)(\frac{1}{2} - \delta)/\delta$	$(1 - \delta \leq x \leq 1)$

For both A and B, $F(x) \equiv G(x) \equiv 0$ for $x < 0$ and $F(x) \equiv G(x) \equiv 1$ for $x > 1$. For B, it will be shown that there exist values of n greater than any preassigned arbitrarily large number, such that the probability of rejecting the hypothesis when it is false is less than $\beta_1 + \beta_2 + \epsilon$ where ϵ is an arbitrarily small positive number.

Let h_1, h_2, h_3 denote the number of observations on X which fall in the intervals $0 < x \leq \frac{1}{2} - \delta, \frac{1}{2} - \delta < x \leq \frac{1}{2} + \delta, \frac{1}{2} + \delta < x \leq 1$ respectively for a fixed value of n . Let h'_1, h'_2, h'_3 be the corresponding numbers for Y . For a fixed n , the probability of a set $h_1, h_2, h_3, h'_1, h'_2, h'_3$ is the same whether the samples be drawn from A or B. From (4), (7), and the stochastic convergence of $m/2n$, it follows that we can find an N such that for all $n > N$ the probability is at

least $1 - \epsilon/2$ of the occurrence of a set $h_1, h_2, h_3, h'_1, h'_2, h'_3$ for which $y_{m_1}, x_{n+1}, y_{m_2}$ will fall in the interval $(\frac{1}{2} - \delta, \frac{1}{2} + \delta)$. Furthermore, for fixed h_2, h'_2 the distribution within the interval is the same whether the sample came from A or B. Hence, even when the sample is drawn from B, for n sufficiently large,

$$P\{m > m_1\} < \beta_1 + \frac{\epsilon}{2}$$

$$P\{m < m_2\} < \beta_2 + \frac{\epsilon}{2}.$$

That is, for samples of sufficiently large size from B, the probability of rejecting the null hypothesis is at most $\beta_1 + \beta_2 + \epsilon$. Since $\beta_1 + \beta_2 < 1$ and ϵ is arbitrarily small, the probability can be made less than one and the test is not consistent in the case of two intervals.

In the case of four intervals, the proof is similar. In this case, we assume that the second sample has size $4n$. Clearly, $n_1/4n, n_2/4n, n_3/4n$, and $n_4/4n$ converge stochastically to $\frac{1}{4}$. If C^* is the significance point defined by (3), the sequence

$$C^*(n) \qquad n = 1, 2, \dots$$

converges to zero. Now consider two pairs, A and B, of populations. A is the same as before and B consists of one uniform distribution and one which is identical with the uniform distribution in small intervals containing $x = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$, and 1, but is different everywhere else. As before, $F(x) \equiv G(x) \equiv 0$ for $x < 0$ and $F(x) \equiv G(x) \equiv 1$ for $x > 1$. Then for B, when n is large, the behavior of C , except for a probability arbitrarily near zero, will depend only on the intervals of coincidence. Hence for B

$$P\{C > C^*\} < \beta' + \epsilon$$

where ϵ is any arbitrarily small positive quantity.

Returning to the case of two intervals, if the samples are from different populations and if their cumulative distribution functions are identical in the neighborhood of their medians, the test is not consistent. If such a possibility is excluded from the class of admissible alternatives, we may expect that the test will be consistent. For example, if the class of alternatives is limited to those where $G(x) \equiv F(x + c)$, c a constant, the test will be consistent. A similar remark holds for the case of four intervals or for any fixed finite number of intervals. It appears, however, that if the number of intervals is a function of the sample size (say \sqrt{n}) and becomes infinite with sample size, a test of this kind will be consistent with respect to a general class of alternatives.

REFERENCES

- [1] A. WALD AND J. WOLFOWITZ, "On a test whether two samples are from the same population," *Annals of Math. Stat.*, Vol. XI (1940), pp. 147-62.
- [2] H. C. MATHESEN, "A method of testing the hypothesis that two samples are from the same population," *Annals of Math. Stat.*, Vol. XIV (1943), pp. 188-94.

HENRY LEWIS RIETZ—IN MEMORIAM

BY A. R. CRATHORNE

University of Illinois

Forty odd years ago few if any American college catalogs mentioned the words "mathematical statistics." The word "actuary" often called for the use of a dictionary. Some courses in the theory of probability, theory of errors, or method of least squares touched on some phases of statistics but aside from this there was little interest in the subject. In England at this time Karl Pearson was well started in his work at University College but "Student" was an undergraduate student. In Germany, Lexis was finishing his somewhat unrecognized labors at Goettingen. In Denmark, Thiele, and in Norway, Charlier were lecturing and writing on statistics from their own individual viewpoints.

During the four decades which have passed, the interest in theoretical statistics in the United States has increased to the point where it has a well established journal of its own and few university mathematical departments fail to list statistical courses. In this growth no one has had more influence than the subject of this memoir. His published papers, his personality, his students and his well directed energy have all been more than helpful in putting mathematical statistics where it is today.

Henry Lewis Rietz, son of Jacob and Tabitha Jane Rietz, was born August 24, 1875 at Gilmore, Ohio. He attended the local schools and in 1895 entered Ohio State University receiving his B.S. degree in 1899. After graduation he went to Cornell University as scholar, then fellow and assistant in mathematics. During his stay at Cornell he was closely associated with two other mathematical students, J. W. Young and H. W. Kuhn, later heads of the departments of mathematics at Dartmouth and Ohio State University respectively. In his last year Rietz was particularly interested in group theory and worked for his doctorate with Professor G. A. Miller who was then a member of Cornell's faculty. His dissertation was "On primitive groups of odd orders," published later in the *American Journal of Mathematics* and referred to in the *Encyclopedie des Sciences Mathematiques*. After receiving the Ph.D. in 1902 he spent one year as professor of mathematics and astronomy at Butler College in Indianapolis.

In 1903, Rietz accepted an instructorship at the University of Illinois where he stayed until 1918 becoming full professor in the meantime. In 1918 he was called to the University of Iowa as head of the department of mathematics, a position he held until his retirement in 1942.

During his first year at Illinois his interests were mainly in pure mathematics. His advanced courses were "Theory of Invariants" during the first semester and "Higher Plane Curves" during the second. During the next year a demand arose for some course in statistics. None of the members of the mathematics department were particularly prepared to give such a course but Rietz was induced to try it. The result was that he offered a course "Averages and

Mathematics of Investment" This curious title was evidence of the fact that actuarial science had not reached the independent position that it has at the present time In the following year he was appointed to the position of statistician of the College of Agriculture and from that time on during his stay at Illinois he divided his time equally between the department of mathematics and that college. His work as statistician was mainly supervision of the statistical work in the published bulletins. The first publication in the statistical field under his name was the 32-page appendix to Dean Davenport's treatise on breeding.

The first published statistical study was in 1908, a master's thesis for Miss Shade on "Correlation of efficiency in mathematics and in other subjects," printed as one of a series of University Studies It is interesting to recall the attention which this paper received, especially from educational circles. It seemed to fix the method and form of calculation of correlation coefficients which occupied the time of many people during the following years. In these early years it was rather difficult to find a place of publication for a mathematical paper on statistics. Mathematical journals were somewhat reluctant in accepting articles. I remember one occasion when he jokingly complained of the correspondence necessary to explain to an editor the word "correlation" used in a paper.

From 1908 on, Rietz published a long list of papers on statistical topics, some purely theoretical, some expository, some arising out of his connection with the college of Agriculture. Together with his later actuarial studies the list totals 150 titles, the more important of which are included in this article. His much quoted paper of 1920 on "Urn Schemata" pleased him more than any other paper His little book *Mathematical Statistics*, one of the Carus mathematical monographs, written in 1926 was the basis for many university courses for years afterward.

In 1909 the American Institute of Actuaries was organized in Chicago, and Rietz was a charter member. He took particular delight in this organization and was rarely absent from the meetings. He was elected vice-president in 1919 He liked meeting practical actuaries and had a wide acquaintance among them In 1916 he was appointed a member of the Illinois Pension Laws Commission and became its actuary From that time on his interests were pretty evenly divided between mathematical statistics and actuarial problems connected with pensions. He was appointed actuary of the Chicago Pension Commission in 1926, was consulting actuary for the Presidents National Committee on Economic Security 1934, and was a member of the board of trustees of the Teacher's Insurance and Annuity Association 1934-38 His services as a consulting actuary were sought by a great many pension projects both in educational and in business circles. When he went to the University of Iowa in 1918, he accentuated actuarial theory in his teaching. Under his leadership the department became an outstanding school in this field Many of his students hold prominent positions in the actuarial world

In 1923, Rietz with eight others was appointed a member of the Committee on the Mathematical Analysis of Statistics of the division of Physical Sciences of the National Research Council. The work of this committee developed into the preparation of the "Handbook of Mathematical Statistics" with Rietz as Editor-in-Chief. This work had considerable use for a number of years after its publication and is an important part of the history of mathematical statistics in this country. A Russian edition of this book appeared in 1927 with a very long preface as a sort of apology for the translation. A few excerpts from this preface are—"Mathematical statistics is a purely technical weapon, politically unbiased, can serve with equal facility either to thwart or to expedite the movement for the emancipation of the proletariat depending in whose hands it happens to be;" "Mathematical statistics has nothing to do with philosophical enlightenment;" "Hence this book harbors no dangers for a soviet reader;" "The fact that the western authors work in a bourgeois society has no bearing on their methods."

The Institute of Mathematical Statistics was organized in 1935 with Rietz as the informal chairman of a steering committee during the months of discussion preceding the organization. He thus became the logical first president. He has taken a more than active interest in the Institute,—as a contributor to the *Annals*, as one of its editors, as general counselor, as a good friend. In appreciation of this and in recognition of his contributions to the initiation and development of mathematical statistics in America, the 1943 volume of the *Annals of Mathematical Statistics* was dedicated to him, on the occasion of his retirement after twenty five years of service as head of the department of mathematics at the University of Iowa.

Professor Rietz received many honors in other fields. He was President of the Mathematical Association of America in 1924, vice-president of the American Statistical Association in 1925, vice-president of the American Mathematical Society 1928-9, and a member of the editorial staffs of the *Bulletin* and the *Transactions* of that society for many years, president of the Iowa Academy of Science, 1931. He was starred in American Men of Science, a fellow of the Royal Statistical Society of London, and of the American Association for the Advancement of Science. He took great interest in local affairs and held many offices in church, social and business organizations in Iowa City. His mind was not altogether centered upon research and the development of mathematics and statistics. He took great pride in his teaching. He was the principal author of a number of college texts in mathematics that had wide use, and was on many committees concerned with the problem of teaching mathematics to undergraduates.

At the time of his retirement, Professor Rietz was in failing health and was practically an invalid until the time of his death at the University Hospital at Iowa City on December 7, 1943. He leaves a brother, Professor John Rietz, Morgantown, West Virginia, and a sister, Mrs. T. S. Taylor, Caldwell, New Jersey.

SELECTED TITLES OF PUBLICATIONS BY H. L. RIETZ

1. "On primitive groups of odd order," *Amer Jour of Math*, Vol 26 (1904), pp. 1-30.
2. "On groups in which certain commutative operations are conjugate," *Trans. Amer. Math. Soc.*, Vol 5(1904), pp. 500-508.
3. "Simply transitive groups which are simple groups," *Bulletin Amer. Math. Soc.*, Vol. 11(1905), pp. 545-48.
4. "Statistical Methods. Appendix to Principles of Breeding," *A Treatise on Thrematology* by E. Davenport, Ginn and Co., Boston, 1907, pp 681-713.
5. "Correlation of efficiency in mathematics and efficiency in other subjects—A Statistical Study," Rietz and Shade, The University Studies, Urbana, Illinois, November 1908, 20 pp.
6. *Statistical Methods Applied to the Study of Type and Variability in Corn*, Eugene Davenport and Henry L. Rietz, Bulletin No. 119, Illinois Agricultural Experiment Station, 1907.
7. "On inheritance in the production of butter fat," *Biometrika*, London, Vol. 7(1909), pp. 106-126.
8. "On a mean difference problem that occurs in statistics," *Amer Math. Month*, Vol. 17(1910), pp. 235-40.
9. "On the metabolism experiment as a statistical problem," Rietz and Mitchell. *Jour. of Biol. Chem.*, 1910.
10. *On the Measurement of Correlation with Special Reference to Some Characters of Indian Corn*, Henry L. Rietz and Louis H. Smith, Bulletin No. 148, University of Illinois Agricultural Experiment Station, November 1910.
11. "On the construction and graduation of a rural life table," Rietz, H. L. and Forsyth, C. H., *Record Amer. Inst. of Actuar*, Vol 1(1911), pp 9-19.
12. "On the theory of correlation with special reference to certain significant loci on the plane of distribution in the case of normal correlation," *Annals of Math.*, Vol. 13(1912), pp. 187-199.
13. "Note on the definition of an asymptote," *Amer Math. Month.*, Vol. 19 (1912), pp 89-90.
14. "The determination of the relative volumes of the components of rocks by mensuration methods," Lincoln and Rietz. *Economic Geology*, Vol VIII, No 2, March, 1913.
15. "On the mathematical theory of risk and Landre's theory of the maximum," *Record Amer. Inst. Actuar.*, Vol. 11(1913), pp. 1-14.
16. "On the status of certain current pension funds," *Record Amer. Inst. Actuar.*, Vol. 3(1914), pp. 33-53.
17. "Group Insurance," *Record Amer. Inst. Actuar.*, Vol. 3(1914), pp. 277-79.
18. "Degrees of resemblance of parents and offspring with respect to birth as twins for registered Shropshire sheep" (Rietz, H. L. and Roberts, Elmer.) *Jour. Agric. Res*, Vol. 4, No 6, pp. 479-510.
19. "Note on double interpolation by finite differences," *Record Amer. Inst. Actuar.*, Vol. 4(1915), pp. 15-22
20. "On the correlation of marks in mathematics and law," *Jour. Educ. Psych.*, Vol. 7, No. 2, pp 87-92.
21. "The operation of pension laws in foreign countries," Report of Illinois Pension Laws Commission, 1916, pp. 19-35.
22. "Actuarial Report on Pension Funds for Public Employees of Illinois," H. L. Rietz, and D. F. Campbell, Report of Illinois Pension Laws Commission, 1916, pp. 72-199.
23. "The underlying principles of a pension plan," H. L. Rietz, G. E. Hooker, and D. F. Campbell, Report of Illinois Pension Laws Commission, 1916, pp. 272-284.
24. "Report of Illinois Pension Laws Commission," Rietz and others, 1916, pp. 310.

25. "On the value of certain proposed refunds payable at the death of an annuitant under a pension system," *Record Amer. Inst. Actuar.*, Vol. 6(1917), pp. 62-75.
26. "A statistical study of some indirect effects of certain selections in the breeding of indian corn," Rietz, H. L. and Smith, L. H., *Jour. of Agric. Res.* Vol. 11(1917), pp. 105-48.
27. "Methods of providing for expenses of new business by life insurance companies," *Amer. Econ. Rev.*, Vol. 7(1917), pp. 832-38.
28. "A Report to the Trustees of the Carnegie Foundation by the Commission Chosen to study and report upon the Proposed Plan of Insurance and Annuities," Rietz and others, 19 pp.
29. "Pensions for public employees," *Amer. Polit. Sci. Review*, Vol. 12(1918), pp. 265-68.
30. "Statistical methods for preparation for war department service," *Amer. Math. Month.*, Vol. 26(1919), pp. 99-100.
31. "Recent developments in pension plans for public employees," *Record Amer. Inst. Actuar.*, Vol. 8(1919), pp. 1-12.
32. "Scope and advantages of courses of instruction on life insurance in American colleges and universities," *Record Amer. Inst. Actuar.*, Vol. 8(1919), pp. 202-06.
33. "The effect of present inflated prices on the future interest rate," *Record Amer. Inst. Actuar.*, Vol. 8(1919), pp. 308-14.
34. "On functional relations for which the coefficient of correlation is zero," *Quart. Pub. Amer. Stat. Assn.*, September 1919, pp. 472-76.
35. "Exposition of the main provisions of the standard plan for a combined comprehensive annuity and insurance system for public employees," Illinois Pension Laws Commission Report, 1919, pp. 25-48.
36. "Recent developments in pension legislation in other states of the United States," Illinois Pension Laws Commission Report, 1919, pp. 201-06.
37. "The world's experience in the operation of public service pension systems," Illinois Pension Commission Report, 1919, p. 207-14.
38. "Summary of the report of the Illinois Pension Laws Commission of 1916," Illinois Pension Commission Report, 1919, pp. 215-22.
39. "Effects of the pension legislation by the fiftieth general assembly of Illinois," Illinois Pension Commission Report, 1919, pp. 223-34.
40. "Illinois state teachers' pension and retirement system," Illinois Pension Commission Report, 1919, pp. 235-38.
41. "Industrial and institutional pension systems," Illinois Pension Commission Report, 1919, pp. 239-50.
42. "Urn schemata as a basis for the development of correlation theory," *Annals of Math.*, Vol. 21(1920), pp. 308-22.
43. "On certain properties of Makeham's laws of mortality," *Amer. Math. Month.*, Vol. 27(1921), pp. 152-65.
44. "Pension systems for insurance company employees," *Record Amer. Inst. Actuar.*, Vol. 10(1921), pp. 1-14.
45. "An elementary exposition of the theorem of Bernoulli with applications to statistics," *Math. Teacher*, Vol. 14(1921), pp. 427-34.
46. "Frequency distributions obtained by certain transformations of normally distributed variates," *Annals Math.*, Vol. 22(1922), pp. 202-300.
47. "Statistics" in a Mathematical Encyclopedia Dictionary," *Amer. Math. Month.*, Vol. 29(1922), pp. 333-337.
48. "On the subject matter of a course in mathematical statistics," Presented before the Mathematical Association of America at the Symposium held in Cambridge, Dec. 29, 1922. *Amer. Math. Month.*, Vol. 30(1923), pp. 155-166.
49. "On certain topics in the mathematical theory of statistics," symposium Lectures before American Mathematical Society, *Bull. Amer. Math. Soc.*, Vol. 30(1924), pp. 417-453.

50. "On annuity rates," *Record Amer. Inst. Actuar.*, Vol. 13(1924), pp. 120-122.
51. "On certain applications of mathematical statistics to actuarial data," *Record Amer. Inst. Actuar.*, Vol. 13(1924), pp. 214-250.
52. "On a certain law of probability of Laplace," *Proc. Internat. Math. Cong.*, Toronto, Vol. 2(1924), pp. 795-799.
53. "Note on average numbers of brothers and of sisters of the boys in families of N children," *Science*, Vol. 60(1924), pp. 46-47.
54. "On the representation of a certain fundamental law of probability," *Trans. Amer. Math. Soc.*, Vol. 27(1925), pp. 197-212.
55. "On applications of statistical methods in actuarial science," *Record Amer. Inst. Actuar.*, Vol. 14(1925), pp. 102-04.
56. "Mathematical background for the study of statistics," H. L. Rietz and A. R. Craithorne, *Jour. Amer. Stat. Assn.*, Vol. 21(1926), pp. 435-440.
57. "On certain applications of the differential and integral calculus in actuarial science," *Amer. Math. Month.*, Vol. 33(1926), pp. 9-23.
58. "Note on the most probable number of deaths," *Record Amer. Inst. Actuar.*, Vol. 16(1927), pp. 25-29.
59. "On certain properties of frequency distributions of the powers and roots of the variates of a given distribution," *Proc. Nat. Acad. Sci.*, Vol. 13(1927), pp. 817-20.
60. Discussion of "Interpolation with modified coefficients," *Record Amer. Inst. Actuar.*, Vol. 16(1927), pp. 232-33.
61. "On the risk problem from a mathematical point of view," *Neuvieme Congres International D'Actuaires, Rapports, Tome II*, (1930), pp. 294-306.
62. "Pensions for superannuated employees," Retiring vice-presidential address before Section K AAAS, *Sci. Month.*, March 1930, pp. 224-30.
63. "On certain properties of frequency distributions obtained by a linear fractional transformation of the variates of a given distribution," *Annals of Math. Stat.*, Vol. 2(1931), pp. 38-47.
64. "Note on the distribution of the standard deviation of sets of three variates drawn at random from a rectangular distribution," *Biometrika*, Vol. 23(1931), pp. 424-26.
65. "Some remarks on mathematical statistics," Retiring president's address before the Iowa Academy of Science, *Science*, Vol. 74(1931), pp. 1-4.
66. "Comments on applications of recently developed theory of small samples," *Jour. Amer. Stat. Assn.*, Vol. 26(1931), pp. 37-44.
67. "A simple non-normal correlation surface," *Biometrika*, Vol. 24(1932), pp. 288-90.
68. "On the Lexis theory and the analysis of variance," *Bull. Amer. Math. Soc.*, Vol. 27(1932), pp. 731-35.
69. "Unemployment and social insurance," *Record Amer. Inst. Actuar.*, Vol. 23(1934), pp. 147-52.
70. "On the frequency distribution of certain ratios," *Annals of Math. Stat.*, Vol. 7(1936), pp. 145-53.
71. "Some topics in sampling theory," *Bull. Amer. Math. Soc.*, Vol. 43(1937), pp. 209-30.
72. "Collective insurance," *Bull. Amer. Assn. Univ. Prof.*, Vol. 23(1937), pp. 278-81.
73. "On the distribution of the 'Student' ratio for small samples from certain non-normal distributions," *Annals Math. Stat.*, Vol. 10(1939), pp. 265-74.
74. "On a recent advance in statistical inference," *Amer. Math. Month.*, Vol. 45(1938), pp. 149-58.

DOCTORATE DISSERTATIONS WRITTEN UNDER THE SUPERVISION OF PROFESSOR RIETZ

- Reilly, John Franklin, 1921. On certain generalizations of osculatory interpolation.
- Weida, Frank M., 1923. The valuation of life annuities with refund of an arbitrarily assigned part of the purchase price.
- Smith, Clarence De Witt, 1928. On generalized Tchebycheff inequalities in mathematical statistics.

- Meyer, Herbert A.*, 1929. On certain inequalities with applications in actuarial science
- Craig, Allen Thornton*, 1931. On the distribution of certain statistics derived from small random samples
- Wilks, Samuel Stanley*, 1931. On the distribution of statistics in samples from a normal population of two variables with matched sampling of one variable
- Fischer, Carl H.*, 1932. On correlation surfaces of sums with a certain number of random elements in common.
- Harper, Floyd S.*, 1935. An actuarial study of infant mortality.
- Oliver, Arthur*, 1935. On certain mathematical developments underlying an analysis of general death rates.
- Knowler, Lloyd A.*, 1937. Actuarial aspects of recent old age security legislation
- Olshen, Abraham C.*, 1937. Transformations of the Pearson type III distribution.
- Berg, William D.*, 1941. Theorems on certain type A difference equation graduations.
- Satterthwaite, Franklin*, 1941. Developments on the theory of Chi-square.
- Garfin, Louis*, 1942. A comparative study of the underlying principles of certain pension schemes for a staff of employees with special reference to teachers and public employees.

The last two dissertations were under the joint supervision of Professor Rietz and Professor A. T. Craig.

BOOKS

1. *College Algebra*, H. L. Rietz and A. R. Crathorne, First edition, 1909. Fourth edition, 1939. Henry Holt and Company, New York.
2. *School Algebra*, two volumes. H. L. Rietz, A. R. Crathorne, E. H. Taylor, 1915. Henry Holt and Company.
3. *Mathematics of Finance*, H. L. Rietz, A. R. Crathorne, J. C. Rietz. First edition 1921, second edition 1929. Henry Holt and Company.
4. *Introductory College Algebra*. H. L. Rietz and A. R. Crathorne. First edition 1923, second edition 1933. Henry Holt and Company.
5. *Handbook of Mathematical Statistics*, H. L. Rietz, Editor-in-chief, 1924, Houghton Mifflin Company.
6. *Mathematical Statistics*. Third Carus Mathematical Monograph, published for the Mathematical Association of America by the Open Court Publishing Co 1927.
7. *Review of Pre-college Mathematics*, C. J. Lapp, F. B. Knight, H. L. Rietz, 1934. Scott, Foresman and Company.
8. *Plane Trigonometry*, H. L. Rietz, J. F. Reilly, Roscoe Woods, 1935. The Macmillan Company.
9. *Plane and Spherical Trigonometry*. H. L. Rietz, J. F. Reilly, Roscoe Woods, 1936. The Macmillan Company.
10. *Intermediate Algebra*, H. L. Rietz, A. R. Crathorne, L. J. Adams, 1942. Henry Holt and Company.
11. *Review of Mathematics for College Students*. C. J. Lapp, F. B. Knight, H. L. Rietz, 1942. Foresman and Company.

NEWS AND NOTICES

*Readers are invited to submit to the Secretary of the Institute
news items of general interest*

Personal Items

Lt. Col. Joseph Berkson is now stationed at Headquarters, Army Air Forces, Air Surgeon's Office, Washington 25, D. C.

Dr. Ernest E. Blanche is now Statistical Director, Office of the Director of Engineering, with the Curtiss-Wright Corporation at Buffalo.

Dr. Alva E. Brandt is overseas serving as Operations Analyst for the Army Air Forces.

Professor W. G. Cochran and Dr. A. M. Mood are serving as Research Mathematicians on a war research project at Princeton University. Professor Cochran is on leave of absence from Iowa State College; Dr. Mood from the University of Texas.

Mr. Robert Dorfman is overseas serving as Operations Analyst with the 13th U. S. Air Force.

Mr. R. M. Foster of Bell Telephone Laboratories has been appointed professor and head of the department of mathematics of the Polytechnic Institute of Brooklyn.

Dr. Andrew I. Peterson is now Director of Manufacturing Research at the Victor Division of the Radio Corporation of America, Camden, N. J.

Dr. Edward Helly, visiting lecturer at the Illinois Institute of Technology, died on November 28, 1943.

Professor Henry L. Rietz died on December 7, 1943 after a long period of illness. An account of Professor Rietz' scientific life and achievements by Professor A. R. Crathorne appears on pp. 102-108 of the present issue of the *Annals*.

New Members

The following persons have been elected to membership in the Institute:

Allen, Roy George Douglas. D Sc. (London) Reader in Economic Statistics, University of London. Apt. 219, 2745 20th St. NW, Washington 8, D. C.

Burk, Mrs. Marjorie F. A.B. (Hunter) Assoc. Statistician, Meteorology, HQ, AAF. 1912 Third St NE, Washington 2, D. C.

Churchman, C. West. Ph D (Pennsylvania) Research Statistician, Frankford Arsenal; and Lecturer in Philosophy, Univ. of Pennsylvania. Frankford Arsenal, Philadelphia, Pa.

Crump, S. Lee. B.S. (Cornell) Research Associate, Iowa State College. Statistical Lab, Iowa State College, Ames, Iowa.

Divatia, M. V. M.A. (Columbia) Statistical Officer, Dept. of Industries and Civil Supplies, New Delhi, India.

Freund, John E. B.A. (U.C.L.A.) Box 4221 Westwood Village Station, Los Angeles 24, Calif.

- Gill, John P. Statistician, Dept. of Research and Statistics, Federal Reserve Bank, Dallas, Texas.
- Lindsey, Fred D. M.A. (George Washington Univ.) 628 West 114 St., New York, N. Y.
- Maloney, Clifford J. M.A. (Minnesota) 2d Lt., Sig. C. 153^d 14th St. N., Arlington, Va.
- Mandel, John. Licence en Sciences (Univ. of Brussels) 45 Kew Gardens Rd., Kew Gardens, N. Y.
- Mathleus, George John. B.L. (Univ. of Dayton) Shop followup - Assembly Planning, Douglas, Long Beach. 1853 Poppy St., N. Long Beach, Calif.
- McIntyre, Francis E. Ph.D. (Chicago) Program Officer, Foreign Economic Administration Rm. 2446 Temporary U Bldg., Washington 25, D. C.
- Raybould, Ethel H. M.A. (Queensland) Lecturer in Mathematics. The University of Queensland, Brisbane, Australia.
- Rosenblatt, Alfred. Ph.D. (Cracow) Catedratico de la Universidad de San Marcos, Lima. Calle Atahualpa 192, Miraflores, Peru.
- Saunders, Robert J. B.S. (Mass. Inst. of Tech.) Captain, Ordnance Dept., Inspection Sec., Amm'n Branch, Office Chief of Ordnance. Office Field Director Ammunition Plants, 3637 Lindell Blvd., St. Louis 8, Missouri.
- Schwartz, David H. B.S. (C.C.N.Y.) Assoc. Statistician, Office of the Quartermaster General. 338 N. Geo. Mason Dr., Arlington, Va.
- Solomons, Leonard M. B.A. (Columbia) Time Study Man. 150-11 88 Ave., Jamaica 2, N. Y.
- Steinberg, Joseph. B.S. (C.C.N.Y.) Associate Statistician, Bureau of Research and Statistics, Social Security Board. 5041 North Capitol Street, Washington 11, D. C.
- Tomlinson, Malcolm C. W. 3820 Southern Ave., S.E., Washington, D. C.
- Wilson, Edward F. Asst. Engineer, Special Projects. Bldg 650, Research Center, Aberdeen Proving Ground, Md.

Announcements

Washington Meeting of the Institute

There will be a joint sectional meeting* of the Institute of Mathematical Statistics and the American Statistical Association at the Hotel Statler and George Washington University in Washington, D. C., on Saturday and Sunday, May 6-7, 1944.

On Saturday afternoon there will be a session on the *Theory of Statistical Inference* with Professor A. Wald and Lt. J. H. Curtiss as the speakers. On Sunday morning there will be a session on contributed papers and on Sunday afternoon a session will be devoted to *Time Series* with Professor J. L. Doob and Dr. T. Koopmans as the speakers.

Summer Meeting of the Institute

The summer meeting of the Institute will be held in conjunction with the summer meetings of the American Mathematical Society and the Mathematical Association of America, at Wellesley College, Wellesley, Massachusetts, on August 12-14, 1944. Abstracts of contributed papers for this meeting should be sent (in duplicate) to the Secretary of the Institute before July 1, 1944.

* The Washington meeting of the Institute had been announced in the December, 1943 issue of the *Annals* for April 27-28, 1944, but the date has been revised to May 6-7, in order that the meeting may be held jointly with the Association.

ANNUAL REPORT OF THE PRESIDENT OF THE INSTITUTE

During 1943, the second war year in which a regular annual meeting of the Institute was not held, the business of the Institute was necessarily largely conducted by mail. As reported in the March issue of the *Annals*, the Secretary, the Editor of the *Annals*, and I met in Pittsburgh in January for discussion of the Institute's affairs. Since three members of the Board of Directors do not constitute a quorum, certain proposals arising from this informal meeting were submitted by mail to the whole Board for action. At the fall meeting of the Institute held in New Brunswick a quorum of the Board was present, Hotelling, Wald, Wilks, Craig, and action was taken on certain matters. The chief subject for consideration was future meetings of the Institute. It was agreed that again in accordance with the request of the O.D.T. no annual meeting should be planned for 1943. Because of the success of local meetings held in New York City in May and in Washington in June, it was voted to repeat these in 1944. The New Brunswick meeting was also very successful, and it is hoped that we may again hold our fall meeting in conjunction with those of the American Mathematical Society and the Mathematical Association of America in 1944. The desirability of other local meetings in addition to those in New York and Washington was recognized but, so far, I know of no plans for any in 1944.

During the year two local chapters of the Institute were organized in Pittsburgh and Washington in accordance with regulations adopted by the Board of Directors, and these were recognized by the Board. The sponsor for the Pittsburgh group is E. G. Olds, and W. G. Madow has the same responsibility for the Washington chapter.

The Membership Committee for 1943 consisted of Vice-President Deming, Chairman, W. G. Cochran, P. S. Dwyer, and A. J. Lotka. As the result of their recommendation the following eighteen new Fellows of the Institute were elected by the Board of Directors: A. H. Copeland, J. H. Curtiss, J. F. Daly, C. E. Dieulefait, H. F. Dodge, Churchill Eisenhart, Will Feller, Milton Friedman, M. A. Girschick, M. H. Hansen, P. G. Hoel, Tjalling Koopmans, A. M. Mood, L. J. Reed, L. E. Simon, F. F. Stephan, W. R. Thompson, and Jacob Wolfowitz.

W. D. Baten, L. A. Aioian, I. W. Burr, and H. F. Dodge, at the request of the Board, continued to serve as a committee for securing additional library subscriptions to the *Annals*. Programs for the New York and New Brunswick meetings were in charge of Vice-President Wald, and that for the Washington meeting was arranged by W. G. Madow.

A proposal originating with Harold Hotelling that the Institute consider petitioning the Federal Government that the W. P. A. Computing Project be made a permanent computing group for the service of scientific research in the construction of important numerical tables, was referred to a committee consisting of A. R. Crathorne, Chairman, P. S. Dwyer, and Will Feller. As a result of their report, the Board appointed the following permanent committee on tabular computation, P. S. Dwyer, Chairman, Churchill Eisenhart, and Will Feller. It

is expected that this committee will cooperate with representatives of other scientific organizations interested in tabular computation.

The Nominating Committee for the recently held election of the Institute was A. T. Craig, Chairman, B. H. Camp, and J. H. Curtiss. G. W. Snedecor served the Institute as its representative on the Council of the American Association for the Advancement of Science.

To all of those mentioned the Institute is indebted, and for the Board of Directors I wish to thank them for their special contributions to the Institute.

I have reserved for particular mention the services of Vice-President Deming who was appointed by the Board the official representative of the Institute to deal with officials of the Selective Service Board in Washington relative to the deferment from military service of competent and experienced statisticians engaged in work important for the prosecution of the war. He gave much time and effort to this on his own initiative, and the Institute and all statisticians owe him much in that he has been very successful in convincing important Washington officials of the value of the services being rendered by good statisticians.

In December the Board with deep regret accepted the resignation of E. G. Olds who was completing his third year as Secretary of the Institute. No member of the Institute will need to be reminded of the devotion and efficiency with which he filled his office. A mere comparative inspection of his annual reports would reveal to anyone otherwise uninformed how much the Institute owes to him. During his term of office the membership of the Institute has approximately doubled and the financial position of the Institute has been markedly improved in spite of war conditions. For both of these favorable circumstances he deserves the greater part of the credit. In his present position as Chief Statistical Consultant with the Office of Production Research and Development of the War Production Board in association with Holbrook Working, he is still serving the cause of statistics as well as the war effort. The Institute will greatly benefit by the intensive work these men are doing in educating industry in the uses of statistical methods in the control of quality in production.

The annual election of the Institute just concluded by mail resulted in the election of the following officers for 1944: W. A. Shewhart, President; W. G. Cochran and Will Feller, Vice-Presidents; and P. S. Dwyer, Secretary-Treasurer. Professor Dwyer had been appointed Acting Secretary-Treasurer upon the resignation of Professor Olds. I need not remind the members of the Institute that the year 1944 remains critical for the Institute and that these new officers will need the fullest support.

C. C. CRAIG,
President, 1943.

February 15, 1944.

ANNUAL REPORT OF THE SECRETARY-TREASURER OF THE INSTITUTE

During 1943 three meetings of the Institute were held. On May 29, the Institute met jointly with the American Society of Mechanical Engineers at the Engineering Societies Building in New York City, the program having been arranged by Abraham Wald and A. I. Peterson. A meeting consisting of three evening sessions was held on June 17-19, at George Washington University, Washington, D. C. William G. Madow made the necessary preparations for this meeting. On September 12-13, the Institute held its sixth summer meeting at New Jersey College for Women, Rutgers University, New Brunswick, New Jersey. This meeting was held in conjunction with the summer meetings of the American Mathematical Society and the Mathematical Association of America, and the program was arranged by Abraham Wald. All three of these meetings were well attended.

Early in the year a petition was granted for the establishment of a Pittsburgh Chapter of the Institute. This organization, formerly known as the Society of Quality Control Statisticians, held its first meeting under the auspices of the Institute on June 19, and a second meeting was held on October 9.

As recorded in an addendum to last year's report, Vice-President E. L. Dodd died on January 9, 1943, and Dr. W. E. Deming was appointed to fill out the unexpired term.

Professor H. L. Rietz died on December 7, 1943. A statement of appreciation for his work in mathematical statistics and in connection with the Institute has been prepared by Professor A. R. Crathorne and appears in the present issue of the *Annals*.

The following financial statement covers a period from December 10, 1942, to December 21, 1943 (the books and records of the Treasurer have been audited by Paul S. Dwyer and found to be in agreement with the statement as submitted):

FINANCIAL STATEMENT

December 10, 1942, to December 21, 1943

RECEIPTS

BALANCE ON HAND, December 10, 1942..	\$2,155.13
DUES.....	2,640.62
SUBSCRIPTIONS.....	1,631.67
SALES OF BACK NUMBERS.....	901.47
MISCELLANEOUS.....	6.07
<hr/>	
Total Receipts	\$7,334.96

EXPENDITURES

ANNALS OFFICE	\$5 25
WAVERLY PRESSES	
Printing and Mailing <i>Annals</i> —4 issues	2,763 28
BACK NUMBERS OFFICE	
Purchase of back numbers from H. C. Carver	\$171 56
Reprinting 200 copies of Vol. VII, No. 1	108.08
	279 64
LIBRARY COMMITTEE	36.11
SECRETARY-TREASURER'S OFFICE	
Printing and Supplies	\$135.00
Binding	2.25
Postage	135.77
Clerical Help	171.20
	444.22
PROGRAMS FOR MEETINGS	44 22
BOARD OF DIRECTORS	41.74
MISCELLANEOUS	5.45
	44.22
Total Expenditures	\$3,619 91
BALANCE ON HAND, December 21, 1943	3,715.05
	\$7,334.96

In comparison with the financial condition of the Institute at the end of 1942, the receipts from dues and subscriptions have increased nearly \$700 and the proceeds from sales of back numbers have decreased nearly \$500. In spite of increased prices, it was possible to reduce expenditures by approximately \$800. Thus the Institute finds itself in a somewhat more favorable position than at the end of last year.

EDWIN C. OLDS,
Secretary-Treasurer.

December 27, 1943.

CONSTITUTION OF THE INSTITUTE OF MATHEMATICAL STATISTICS

ARTICLE I

NAME AND PURPOSE

1. This organization shall be known as the Institute of Mathematical Statistics.
2. Its object shall be to promote the interests of mathematical statistics.

ARTICLE II

MEMBERSHIP

1. The membership of the Institute shall consist of Members, Junior Members, Fellows, Honorary Members, and Sustaining Members.

2. Voting members of the Institute shall be (a) the Fellows, and (b) all others, Junior Members excepted, who have been members for twenty-three months prior to the date of voting.

3. No person shall be a Junior Member of the Institute for more than a limited term as determined by the Committee on Membership and approved by the Board of Directors.

ARTICLE III

OFFICERS, BOARD OF DIRECTORS, AND COMMITTEE ON MEMBERSHIP

1. The Officers of the Institute shall be a President, two Vice-Presidents, and a Secretary-Treasurer. The terms of office of the President and Vice-Presidents shall be one year and that of the Secretary-Treasurer three years. Elections shall be by majority ballots at Annual Meetings of the Institute. Voting may be in person or by mail.

(a) Exception. The first group of Officers shall be elected by a majority vote of the individuals present at the organization meeting, and shall serve until December 31, 1936.

2. The Board of Directors of the Institute shall consist of the Officers, the two previous Presidents, and the Editor of the Official Journal of the Institute.

3. The Institute shall have a Committee on Membership composed of three Fellows. At their first meeting subsequent to the adoption of this Constitution, the Board of Directors shall elect three members as Fellows to serve as the Committee on Membership, one member of the Committee for a term of one year, another for a term of two years, and another for a term of three years. Thereafter the Board of Directors shall elect from among the Fellows one member annually at their first meeting after their election for a term of three years. The president shall designate one of the Vice-Presidents as Chairman of this Committee.

ARTICLE IV

MEETINGS

1. A meeting for the presentation and discussion of papers, for the election of Officers, and for the transaction of other business of the Institute shall be held annually at such time as the Board of Directors may designate. Additional meetings may be called from time to time by the Board of Directors and shall be called at any time by the President upon written request from ten Fellows. Notice of the time and place of meeting shall be given to the membership by the Secretary-Treasurer at least thirty days prior to the date set for the meeting. All meetings except executive sessions shall be open to the public. Only papers accepted by a Program Committee appointed by the President may be presented to the Institute.

2. The Board of Directors shall hold a meeting immediately after their election and again immediately before the expiration of their term. Other meetings of the Board may be held from time to time at the call of the President or any two members of the Board. Notice of each meeting of the Board, other than the two regular meetings, together with a statement of the business to be brought before the meeting, must be given to the members of the Board by the Secretary-Treasurer at least five days prior to the date set therefor. Should other business be passed upon, any member of the Board shall have the right to reopen the question at the next meeting.

3. The Committee on Membership shall hold a meeting immediately after the annual meeting of the Institute. Further meetings of the Committee may be held from time to time at the call of the Chairman or any member of the Committee provided notice of such call and the purpose of the meeting is given to the members of the Committee by the Secretary-Treasurer at least five days before the date set therefor. Should other business be passed upon, any member of the Committee shall have the right to reopen the question at the next meeting.

4. At a regularly convened meeting of the Board of Directors, four members shall constitute a quorum. At a regularly convened meeting of the Committee on Membership, two members shall constitute a quorum.

ARTICLE V

PUBLICATIONS

1. The *Annals of Mathematical Statistics* shall be the Official Journal for the Institute. The Editor of the *Annals of Mathematical Statistics* shall be a Fellow appointed by the Board of Directors of the Institute. The term of office of the Editor may be terminated at the discretion of the Board of Directors.

2. Other publications may be originated by the Board of Directors as occasion arises.

ARTICLE VI

EXPULSION OR SUSPENSION

1. Except for non-payment of dues, no one shall be expelled or suspended except by action of the Board of Directors with not more than one negative vote.

ARTICLE VII

AMENDMENTS

1. This constitution may be amended by an affirmative two-thirds vote at any regularly convened meeting of the Institute provided notice of such proposed amendment shall have been sent to each voting member by the Secretary-Treasurer at least thirty days before the date of the meeting at which the proposal is to be acted upon. Voting may be in person or by mail.

BY-LAWS

ARTICLE I

DUTIES OF THE OFFICERS, THE EDITOR, BOARD OF DIRECTORS, AND COMMITTEE ON MEMBERSHIP

1. The President, or in his absence, one of the Vice-Presidents, or in the absence of the President and both Vice-Presidents, a Fellow selected by vote of the Fellows present, shall preside at the meetings of the Institute and of the Board of Directors. At meetings of the Institute, the presiding officer shall vote only in the case of a tie, but at meetings of the Board of Directors he may vote in all cases. At least three months before the date of the annual meeting, the President shall appoint a Nominating Committee of three members. It shall be the duty of the Nominating Committee to make nominations for Officers to be elected at the annual meeting and the Secretary-Treasurer shall notify all voting members at least thirty days before the annual meeting. Additional nominations may be sub-

mitted in writing, if signed by at least ten Fellows of the Institute, up to the time of the meeting.

2. The Secretary-Treasurer shall keep a full and accurate record of the proceedings at the meetings of the Institute and of the Board of Directors, send out calls for said meetings and, with the approval of the President and the Board, carry on the correspondence of the Institute. Subject to the direction of the Board, he shall have charge of the archives and other tangible and intangible property of the Institute, and once a year he shall publish in the *Annals of Mathematical Statistics* a classified list of all Members and Fellows of the Institute. He shall send out calls for annual dues and acknowledge receipt of same; pay all bills approved by the President for expenditures authorized by the Board or the Institute; keep a detailed account of all receipts and expenditures, prepare a financial statement at the end of each year and present an abstract of the same at the annual meeting of the Institute after it has been audited by a Member or Fellow of the Institute appointed by the President as Auditor. The Auditor shall report to the President.

3. Subject to the direction of the Board, the Editor shall be charged with the responsibility for all editorial matters concerning the editing of the *Annals of Mathematical Statistics*. He shall, with the advice and consent of the Board, appoint an Editorial Committee of not less than twelve members to co-operate with him; four for a period of five years, four for a period of three years, and the remaining members for a period of two years, appointments to be made annually as needed. All appointments to the Editorial Committee shall terminate with the appointment of a new Editor. The Editor shall serve as editorial adviser in the publication of all scientific monographs and pamphlets authorized by the Board.

4. The Board of Directors shall have charge of the funds and of the affairs of the Institute, with the exception of those affairs specifically assigned to the President or to the Committee on Membership. The Board shall have authority to fill all vacancies ad interim, occurring among the Officers, Board of Directors, or in any of the Committees. The Board may appoint such other committees as may be required from time to time to carry on the affairs of the Institute.

5. The Committee on Membership shall prepare and make available through the Secretary-Treasurer an announcement indicating the qualifications requisite for the different grades of membership.

ARTICLE II

DUES

Members shall pay five dollars at the time of admission to membership and shall receive the full current volume of the Official Journal. Thereafter, Members shall pay five dollars annual dues. The annual dues of Junior Members shall be two dollars and fifty cents.

The annual dues of Fellows shall be five dollars. The annual dues of Sustaining Members shall be fifty dollars. Honorary Members shall be exempt from all dues.

(a) Exception. In the case that two Members of the Institute are husband and wife and they elect to receive between them only one copy of the Official Journal, the annual dues of each shall be three dollars and seventy-five cents.

2. Annual dues shall be payable on the first day of January of each year.

3. The annual dues of a Fellow, Member, or Junior Member include a subscription to the Official Journal. The annual dues of a Sustaining Member include two subscriptions to the Official Journal.

4. It shall be the duty of the Secretary-Treasurer to notify by mail anyone whose dues

may be six months in arrears, and to accompany such notice by a copy of this Article. If such person fail to pay such dues within three months from the date of mailing such notice, the Secretary-Treasurer shall report the delinquent one to the Board of Directors, by whom the person's name may be stricken from the rolls and all privileges of membership withdrawn. Such person may, however, be re-instated by the Board of Directors upon payment of the arrears of dues.

ARTICLE III

SALARIES

1. The Institute shall not pay a salary to any Officer, Director, or member of any committee.

ARTICLE IV

AMENDMENTS

1. These By-Laws may be amended in the same manner as the Constitution or by a majority vote at any regularly convened meeting of the Institute, if the proposed amendment has been previously approved by the Board of Directors.

•

FURTHER CONTRIBUTIONS TO THE PROBLEM OF SERIAL CORRELATION

BY WILFRID J. DIXON

Princeton University

1. Introduction. Recently, there has been an increasing interest in the study of the serial correlation of observations. The development of the distribution theory and significance criteria was retarded by the fact that the successive differences or successive products of statistical variates are not independent. However, these difficulties have been overcome to a considerable extent by recent work of several authors. In order to indicate the nature of the contributions embodied in the present paper, it will be necessary to describe rather precisely the contributions of these authors

Suppose x_1, x_2, \dots, x_n are n independent observations of a random variable x which is normally distributed with mean \bar{x} and variance σ^2 . Let us define

$$\begin{aligned} \delta_{n-1}^2 &= \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 & \delta_n^2 &= \sum_{i=1}^n (x_{i+1} - x_i)^2 \\ (1.1) \quad C_{n-1} &= \sum_{i=1}^{n-1} (x_i - \bar{x})(x_{i+1} - \bar{x}) & C_n &= \sum_{i=1}^n (x_i - \bar{x})(x_{i+1} - \bar{x}) \\ {}_LC_n &= \sum_{i=1}^n (x_i - \bar{x})(x_{i+L} - \bar{x}) & V_n &= \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

in which $x_{n+1} = x_1$. The ratio of any of the first five values to V_n will be a measure of the relation between the successive observations x_i .

Von Neumann [2] has studied the ratio $\eta = \delta_{n-1}^2/V_n$. He obtains an expression for the sampling distribution of the ratio η . He solves the equivalent problem of determining the distribution of $\sum_{i=1}^{n-1} A_i y_i^2$ where the point $(y_1, y_2, \dots, y_{n-1})$ is uniformly distributed over the spherical surface $\sum_{i=1}^{n-1} y_i^2 = 1$ and the A_i are the characteristic values of δ_{n-1}^2 . He obtains the distribution $\omega(\gamma)$ of $\gamma = \sum_{i=1}^m B_i x_i^n$ (m even) where the point (x_1, x_2, \dots, x_n) is uniformly distributed over the spherical surface $\sum_{i=1}^n x_i^2 = 1$ and $B_1 \geq B_2 \geq \dots \geq B_m$. $\omega(\gamma)$ is found by solving the equation

$$(1.2) \quad \int_{B_m}^{B_1} (\gamma - z)^{-1/2} \omega(\gamma) d\gamma = \prod_{i=1}^m (B_i - z)^{-1/2}.$$

The distribution of η is then a special case of this distribution. The first four moments were obtained by Williams [5] by the use of a generating function. In

the present paper we shall study the ratio δ_n^2/V_n . The moments of this ratio will be developed and the moments and approximate distribution of $[2 - \delta_n^2/V_n]^2$.

Von Neumann [4] in a paper which removed a restriction (that m be even) on the distribution of η indicates how to determine the distribution of C_{n-1}/V_n . Koopmans [9] considers the stochastic process $x_t = \rho x_{t-1} + z_t$ ($t = 1, 2, \dots$), $|\rho| < 1$. The z_t are independent drawings from a normal distribution with zero mean and variance σ^2 . To test the hypothesis that $\rho = 0$ he shows that it is sufficient to know the distribution of C_{n-1}/V_n . He finds the distribution of C_{n-1}/V_n and C_n/V_n but finds that the numerical computation of these functions is very cumbersome. This prompts him to obtain approximate formulas for these distributions. The approximate formula for the distribution of $\bar{r} = C_n/V_n$ is

$$(1.3) \quad (\frac{1}{2}n - 1)2^{1/n} \pi^{-1/2} \int_0^{\arccos \bar{r}} (\cos \alpha - \bar{r})^{1/n-2} \sin \frac{1}{2}n\alpha \sin \alpha d\alpha.$$

A similar approximation will be used in this paper to find the moments of C_n/V_n . It will be shown how good the approximation is and how by using this approximation we may obtain a tabled function (Pearson Type I) which fits the distribution of $1 - (C_n/V_n)^2$ up to $\frac{1}{2}n$ moments.

The quantity $1 - (C_n/V_n)^2$, we shall find, is equivalent to a likelihood ratio function for testing the hypothesis that the serial correlation is zero.

Anderson [8] obtained the distribution of ${}_L C_n/V_n = {}_L R_n$. He proved that the distribution of ${}_L R_n$ is the same as that of ${}_1 R_n$ when L and n are prime to each other. He has computed the 1 per cent and 5 per cent significance values ($L = 1$) up to $n = 75$. For values of $n > 75$ he indicates that a normal distribution which is an asymptotic approximation may be used. He has also computed some significance values for the cases of $N/L = 2, 3, 4$.

In this paper we shall develop the moments of ${}_L R_n$.

The use of the ratio η in the study of serial effects in ballistics at Aberdeen Proving Ground is given in references [1] and [2]. The use of the ratios C_{n-1}/V_n and C_n/V_n in the study of economic time series is discussed by Koopmans [9].

2. Likelihood criteria. Given a sample of n observations x_1, x_2, \dots, x_n we shall assume that they are distributed according to the law:

$$(2.1) \quad dP_n = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{a=1}^n (x_a - a - bx_{a-1})^2} dx_1 \cdots dx_n, \quad (1 \leq l \leq n).$$

It will be convenient to use the phraseology that "the variate at the time α has as its mean value a linear function of the variate at the time $\alpha - l$." We shall take $x_\alpha = x_{\alpha+l}$. Due to the symmetry we may use $\alpha + l$ in place of $\alpha - l$. This will be done to obtain agreement with previous work. We wish to test the hypothesis H_1 that each variate is independent of the other variates, that is, that $b = 0$. The Neyman-Pearson specification of H_1 may be written as follows,

where Ω is the space of admissible values of σ^2 , a and b , and ω the subspace defining H_1 :

$$(2.2) \quad \begin{cases} \Omega: \sigma^2 > 0 & -\infty < a, \quad b < \infty \\ \omega: \sigma^2 > 0 & -\infty < a < \infty, \quad b = 0. \end{cases}$$

The likelihood criterion λ_1 suitable to this hypothesis is the ratio of the maximum (ω (max.)) of (2.1) with the restriction that $b = 0$ to the maximum (Ω (max.)) of (2.1) without this condition. Now,

$$(2.3) \quad \lambda_1 = \frac{dP_n(\omega \text{ max})}{dP_n(\Omega \text{ max})}.$$

We see that the likelihood function is

$$(2.4) \quad L = -n \log (\sqrt{2\pi} \sigma) - \frac{1}{2\sigma^2} \sum_{\alpha=1}^n (x_\alpha - a - bx_{\alpha+1})^2$$

and to maximize L over the space Ω we compute the following derivatives

$$(2.5) \quad \begin{aligned} \frac{\partial L}{\partial a} &= \frac{1}{\sigma^2} \sum (x_\alpha - a - bx_{\alpha+1}), \\ \frac{\partial L}{\partial b} &= \frac{1}{\sigma^2} \sum (x_\alpha - a - bx_{\alpha+1})x_{\alpha+1}, \\ \frac{\partial L}{\partial \sigma} &= -\frac{n}{\sigma^2} + \frac{1}{\sigma^3} \sum (x_\alpha - a - bx_{\alpha+1}). \end{aligned}$$

The solutions of the equations obtained by setting the above derivatives equal to zero are:

$$(2.6) \quad \begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \sum (x_\alpha - \hat{a} - \hat{b}x_{\alpha+1})^2 \\ \hat{a} &= \frac{1}{n} \sum x_\alpha (1 - \hat{b}) \\ \hat{b} &= \frac{n \sum x_\alpha x_{\alpha+1} - (\sum x_\alpha)^2}{n \sum x_\alpha^2 - (\sum x_\alpha)^2}. \end{aligned}$$

If we now maximize L over the space ω we obtain

$$(2.7) \quad \begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \sum (x_\alpha - \hat{a})^2 \\ \hat{a} &= \frac{1}{n} \sum x_\alpha \end{aligned}$$

so that we will have

$$(2.8) \quad dP_n(\Omega \text{ max}) = [2\pi(1 - \hat{b}^2) \sum (x_\alpha - \bar{x})^2]^{-\frac{1}{2}n} e^{-\frac{1}{2}n},$$

$$(2.9) \quad dP_n(\omega \text{ max}) = [2\pi \sum (x_\alpha - \bar{x})^2]^{-\frac{1}{2}n} e^{-\frac{1}{2}n},$$

$$(2.10) \quad \lambda_1 = (1 - \hat{b}^2)^{\frac{1}{2}n},$$

where b is defined as in (2.6) above. If we set $a = 0$ in (2.1) we may follow a similar procedure and find the criterion ${}_0\lambda_1 = (1 - \hat{b}_0^2)^{1/2}$ for testing the hypothesis that $b = 0$ if it is known that the population mean equal zero. Here

$$(2.11) \quad \hat{b}_0 = \frac{\sum x_a x_{a+l}}{\sum x_a^2}.$$

We notice that \hat{b} is the criterion chosen by R. L. Anderson as a measure of serial correlation. He has obtained the distribution of \hat{b} and has computed a number of significance values from this distribution. The distribution is a function of n and l , and Anderson points out that the larger the values of n the smaller the extent to which the significance values depend on l .

In the next section we shall find a distribution which approximates very closely the exact distribution of \hat{b} and which is independent of the lag l .

3. Moments of the likelihood criteria. We shall determine the moments of \hat{b}_0 and ${}_0\lambda_1$ when the hypothesis H_0 is true and the moments of \hat{b} and λ_1 when the hypothesis H_1 is true. Let us first consider $\hat{b}_0 = \sum x_a x_{a+l} / \sum x_a^2$, the criterion we obtained for testing the hypothesis H_0 . The moment generating function for the joint distribution of $C_0 = \sum x_a x_{a+l} / \sigma^2$ and $V_0 = \sum x_a^2 / \sigma^2$ is

$$(3.1) \quad \begin{aligned} \varphi(t_1, t_2) &= E[\exp(C_0 t_1 + V_0 t_2)] \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp \left(C_0 t_1 + V_0 t_2 - \frac{1}{2\sigma^2} \sum x_a^2 \right) \prod_{a=1}^n dx_a. \end{aligned}$$

By reference to this last expression it can be seen that

$$(3.2) \quad E(\hat{b}_0) = \int_{-\infty}^0 \frac{\partial \varphi(t_1, t_2)}{\partial t_1} \Big|_{t_1=0} dt_2,$$

and further

$$(3.3) \quad E(\hat{b}_0^k) = \int_{-\infty}^0 \cdots \int_{-\infty}^0 \frac{\partial^k \varphi(t_1, t_2)}{\partial t_1^k} \Big|_{t_1=0} \prod_{j=1}^k dt_{2j},$$

in which we set $t_2 = \sum_{j=1}^k t_{2j}$.

Now if we write (3.1) as follows:

$$(3.4) \quad \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2} \sum_{i,j} A_{ij} x_i x_j} \prod dx_i,$$

we see that $\varphi(t_1, t_2) = |A_{ij}|^{-1/2}$ and $A_{ij} = A_{i+a, j+l}$; that is, this determinant is a circulant. Let us write $A_{ij} = a_h$ where h equal $j - i + 1$ or $j - i + 1 + n$ taking that subscript which gives $1 \leq h \leq n$, so that we have

$$(3.5) \quad \varphi^{-2}(t_1, t_2) = \begin{vmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_n & a_1 & a_2 & \cdots & a_{n-1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ a_2 & a_3 & a_4 & \cdots & a_1 \end{vmatrix},$$

which expanded by the method of circulants becomes

$$\prod_{k=1}^n \sum_{i=1}^n a_i \omega_k^{i-1},$$

where the ω_k are the n th roots of unity.

First we shall consider \hat{b}_0 (lag 1). Here $a_1 = 1 - 2t_2$, $a_2 = a_n = -t_1$ and a_3 to a_{n-1} equal zero.

$$\begin{aligned} {}_0\varphi_1^{-2}(t_1, t_2) &= \prod_{k=1}^n (1 - 2t_2 - t_1(\omega_k + \omega_k^{-1})) \\ (3.6) \qquad &= \prod_{k=1}^n \left(1 - 2t_2 - 2t_1 \cos \frac{2\pi k}{n} \right). \end{aligned}$$

For lag l , $a_1 = 1 - 2t_2$, $a_{l+1} = a_{n-l+1} = -t_1$ and the remaining a 's = 0.

$$\begin{aligned} {}_0\varphi_1^{-2}(t_1, t_2) &= \prod_{k=1}^n (1 - 2t_2 - t_1(\omega_k^l + \omega_k^{-l})) \\ (3.7) \qquad &= \prod_{k=1}^n \left(1 - 2t_2 - 2t_1 \cos \frac{2\pi lk}{n} \right) \end{aligned}$$

We shall develop an approximation to these functions (3.6) and (3.7) as follows:

$$(3.8) \qquad {}_0\varphi_1(t_1, t_2) = \prod_{k=1}^n (A + B \cos \alpha_k)^{-1} = e^{-\frac{1}{2} \sum_{k=1}^n \log (A + B \cos \alpha_k)},$$

where in this case $A = 1 - 2t_2$, $B = -2t_1$ and $\alpha_k = 2\pi lk/n$. Let us now alter the form of this exponent and replace the sum of a finite number of terms involving α_k by an integral of a continuous variable α .

$$(3.9) \qquad {}_0\varphi_1(t_1, t_2) = \exp \left(\frac{-n}{4\pi l} \frac{2\pi l}{n} \sum_{k=1}^n \log (A + B \cos \alpha_k) \right).$$

Let us write $2\pi l/n = \Delta\alpha_k$, then we shall have

$$(3.10) \qquad {}_0\varphi_1(t_1, t_2) = \exp \left(\frac{-n}{4\pi l} \sum_{k=1}^n \log (A + B \cos \alpha_k) \Delta\alpha_k \right)$$

If we take $B < A$ we see that $A + B \cos \alpha_k$ is never negative; therefore as we let n increase the summation will approach the value of the integral: $\int_0^{2\pi l} \log (A + B \cos \alpha) d\alpha$. Let us then replace this summation by its limiting integral. The resulting function will then be an approximation for n large enough. We shall obtain then

$$(3.11) \qquad {}_0\varphi_1(t_1, t_2) \sim \exp \left(\frac{-n}{4\pi l} \int_0^{2\pi l} \log (A + B \cos \alpha) d\alpha \right) \qquad B < A,$$

from which by the use of the integral¹ we obtain

$$(3.12) \qquad {}_0\varphi_1(t_1, t_2) \sim \exp \left(-\frac{1}{2} n \log \frac{1}{2} (A + \sqrt{A^2 - B^2}) \right)$$

¹ This integral may be verified by differentiation with respect to the parameter a . It may be found as formula 523 in Pierce's "Short Table of Integrals."

So that

$$(3.13) \quad \varphi_1(t_1, t_2) \sim [1/4(A + \sqrt{A^2 - B^2})]^{-1/2}.$$

We now have $\varphi(t_1, t_2)$ represented approximately by a power of a single quantity. The question of how good this approximation is will be discussed in a later paragraph. This is similar to the approximation used by T. Koopmans in the distribution of \hat{b}_0 . He makes the substitution "to obtain what intuitively seems to be in some sense the closest approximation." He approximates $\prod_{i=1}^T (k_i - \bar{k}_i)^{-1}$

where $\bar{k}_i = \cos \frac{2\pi t}{T}$ by the process followed in (3.8)-(3.13) in order to find the distribution given in (1.3).

To obtain the corresponding function for

$$(3.14) \quad \hat{b} = \frac{\sum x_a x_{a+1} - (\sum x_a)^2/n}{\sum x_a^2 - (\sum x_a)^2/n},$$

we follow the same procedure as above for \hat{b}_0 . Here

$$U = [\sum x_a x_{a+1} - (\sum x_a)^2/n]/\sigma^2,$$

$$V = [\sum x_a^2 - (\sum x_a)^2/n]/\sigma^2,$$

and in (3.5) $a_1 = 1 - 2t_2 + 2(t_1 + t_2)/n$; $a_2 = a_n = -t_1 + 2(t_1 + t_2)/n$; and all the other a 's $= 2(t_1 + t_2)/n$ so that the expansion of this circulant becomes

$$(3.15) \quad \varphi_1^{-2}(t_1, t_2) = \prod_{k=1}^n \left\{ [1 - 2t_2 + 2(t_1 + t_2)/n] + [-t_1 + 2(t_1 + t_2)/n](\omega_k + \omega_k^{-1}) \right. \\ \left. + [2(t_1 + t_2)/n] \sum_{i=2}^{n-1} \omega_k^{i-1} \right\},$$

and since

$$\sum_{i=2}^{n-1} \omega_k^{i-1} = \begin{cases} -(\omega_k + \omega_k^{-1} + 1) & k \neq n, \\ n-3 & k = n, \end{cases}$$

we get

$$(3.16) \quad \prod_{k=1}^{n-1} \{1 - 2t_2 - t_1(\omega_k + \omega_k^{-1})\},$$

and for lag l we get

$$(3.17) \quad \prod_{k=1}^{n-1} \{1 - 2t_2 - t_1(\omega_k^l + \omega_k^{-l})\}.$$

These two results are the same as those obtained previously except that the final term, $1 - 2t_1 - 2t_2 = A + B$, of the products is missing. We will then obtain as an approximation to these finite products $\varphi_1^{-2} = \varphi_1^{-2}/(A + B)$ or

$$(3.18) \quad \varphi_1(t_1, t_2) = [\frac{1}{2}(A + \sqrt{A^2 - B^2})]^{-1/2} \sqrt{A + B}$$

where $A = 1 - 2t_2$ and $B = -2t_1$.

A method of finding the moments of \hat{b}_0 and \hat{b} was outlined in (3.1), (3.2) and (3.3) above. If we perform these operations on ${}_0\varphi_1(t_1, t_2)$ as defined in (3.13) we find

$$\begin{aligned}
 {}_0\varphi_1 &= u^{-\frac{1}{2}n} \\
 (3.19) \quad \frac{\partial \varphi}{\partial t_1} &= -\frac{1}{2}nu^{-\frac{1}{2}n-1} \frac{\partial u}{\partial t_1} \\
 \frac{\partial^2 \varphi}{\partial t_1^2} &= -\frac{1}{2}nu^{-\frac{1}{2}n-2} \left[\left(-\frac{n}{2} - 1 \right) \left(\frac{\partial u}{\partial t_1} \right)^2 + u \frac{\partial^2 u}{\partial t_1^2} \right]
 \end{aligned}$$

where $u|_0 = 1 - 2t_2$, $\frac{\partial u}{\partial t_1}|_0 = 0$, $\frac{\partial^2 u}{\partial t_1^2}|_0 = \frac{-2}{1 - 2t_2}$, etc, and the zero subscript indicates that t_1 has been set equal to zero after differentiation. If these values are substituted and the required number of integrations with respect to t_2 are performed, we find the moments of the criterion \hat{b}_0 when ${}_0H_1$ is true

$$\begin{aligned}
 M_1 &= 0 & M_2 &= \frac{1}{n+2} \\
 M_3 &= 0 & M_4 &= \frac{3}{(n+2)(n+4)} \\
 (3.20) \quad M_5 &= 0 & M_6 &= \frac{15}{(n+2)(n+4)(n+6)} \\
 && \text{etc., or}
 \end{aligned}$$

$$M_{2k-1} = 0$$

$$M_{2k} = \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{(n+2)(n+4) \cdots (n+2k)}.$$

M_{2k} may be verified by the use of an expansion of the generating function (3.13) by a method of Laplace [10]. He gives the expansion of $u^{-\frac{1}{2}n}$ where u is given by the equation $u^2 - 2u + e^2 = 0$ as follows:

$$\begin{aligned}
 u^{-\frac{1}{2}n} &= \frac{1}{2^1} + \frac{ue^2}{2^{1+2}} + \frac{i(i+3)e^4}{1 \cdot 2 \cdot 2^{1+4}} + \frac{i(i+4)(i+5)e^6}{1 \cdot 2 \cdot 3 \cdot 2^{1+6}} \\
 &\quad + \cdots + \frac{i(i+k+1) \cdots (i+2k-1)e^{2k}}{(k-1)! \cdot 2^{1+2k}} + \cdots
 \end{aligned}$$

We see that $u = 1 + \sqrt{1 - e^2}$, and if we set $e = t_1/(\frac{1}{2} - t_2)$ and $i = \frac{1}{2}n$. We obtain ${}_0\varphi_1 = u^{-\frac{1}{2}n}$ as a series in the even powers of t_1 . From this we can see that the odd moments are zero and from the form of the coefficients we can verify M_{2k} .

These moments are not contained in the works of the other authors writing on this subject. Although these moments are obtained from an approximate generating function they are, as will be shown later, the exact, not approximate,

moments for $k < n$, for lag 1 and are the exact moments for $k \leq n/\alpha$ for any lag l where α is the largest common factor of n and the lag l .

To obtain the moments of \hat{b} we follow a similar procedure with $\varphi_1 = u^{-1/2}(1 - 2t_1 - 2t_2)^{1/2}$. Differentiating φ_1 the required number of times with respect to t_1 and integrating an equal number of times with respect to t_2 gives the following moments:

$$\begin{aligned}
 M_1 &= \frac{-1}{n-1} & M_2 &= \frac{1}{n+1} \\
 M_3 &= \frac{-3}{(n-1)(n+3)} & M_4 &= \frac{3}{(n+1)(n+3)} \\
 &\vdots & & \\
 M_{2k-1} &= \frac{-1 \cdot 3 \cdot 5 \cdots (2k-1)}{(n-1)(n+3)(n+5) \cdots (n+2k-1)} \\
 M_{2k} &= \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{(n+1)(n+3) \cdots (n+2k-1)}.
 \end{aligned}
 \tag{3.21}$$

Examination of the moments of \hat{b}_0 will show that $\hat{b}_0 = x$ is distributed according to the law

$$K_1(1-x^2)^{1/2(n-1)} = K_1(1+x)^{1/2(n-1)}(1-x)^{1/2(n-1)}$$

up to n moments. This distribution is symmetric and we may wish a normal approximation to this curve. The mean is zero and the variance is $1/(n+2)$. The λ criterion $\lambda_1^{2/n} = (1 - \hat{b}_0^2) = y$ is distributed according to the law

$$K_2(1-y)^{-1/2}y^{1/2(n-1)}$$

up to $\frac{1}{2}n$ moments. Here the mean is $\frac{n+1}{n+2}$ and the variance is $\frac{2(n+1)}{(n+2)^2(n+4)}$. If we inspect the moments of \hat{b} we see that the distribution of $\lambda_1^{2/n} = (1 - \hat{b}^2) = z$ follows the law

$$K_3(1-z)^{-1/2}z^{1/2(n-2)}$$

up to $\frac{1}{2}n$ moments, which is the same as the distribution immediately preceding except that n is replaced by $n-1$. The distributions (3.22), (3.23), and (3.24) are the same for lag l except that the fit is up to n/α , $n/2\alpha$, and $n/2\alpha$ moments respectively where α is the largest common factor of l and n . These restrictions are necessary since the moments as given in (3.20) and (3.21) are obtained from the approximate generating functions (3.13) and (3.18). The exact generating function is given in (4.8) for lag 1 and it is found that the n th or higher derivatives bring contributions from the part of the generating function which was neglected in approximating the generating function. The additional restriction for lag $l \neq 1$ will be seen in the last two paragraphs of section 4. The extra

factor $\frac{1}{2}$ in the second and third case above is due to the fact that only the even moments of (3.20) and (3.21) are used.

We have in (3.23) and (3.24), then, very close approximations to the distributions for the two λ criteria for testing serial effects

The following table is a comparison of the exact and approximate 5 per cent and 1 per cent points for the distribution of \hat{b} . The exact values are taken from the table given by R. L. Anderson. The normal approximation as given by Anderson in his table does not show such close agreement since he used an asymptotic second moment. He indicated that the exact values would have to be used for values of $n < 75$ in place of the values from the normal approximation which he obtained. Here we see that the normal approximation may be used for n somewhat less than 75. The Pearson Type I approximation was obtained by using the first two moments of \hat{b} . The curve obtained is:

$$(3.25) \quad \frac{(1+x)^{p-1}(1-v)^{q-1}}{B(p, q)2^{p+q-1}}$$

in which $p = \frac{(n-1)(n-2)}{2(n-3)}$ and $q = \frac{n(n-1)}{2(n-3)}$

The exact values marked with an asterisk in the table differ slightly from those previously published. They are more precise values from the exact distribution which R. L. Anderson has made available to the author.

Positive tail

<i>N</i>	<i>Exact</i>	5% <i>Type I</i>	<i>Normal</i>	<i>Exact</i>	1% <i>Type I</i>	<i>Normal</i>
5	.253	.317	.281	.297	.527	.501
10	.360	.362	.350	.525	.533	.541
15	.328	.329	.323	.475	.477	.486
20	.299	.299	.296	.432	.433	.440
25	.276	.276	.274	.398	.398	.404
30	.257	.257	.255	.370	.371	.375
45	.218	.218	.217	.313*	.313	.316
75	.174*	.174	.174	.250	.250	.251

Negative tail

<i>N</i>	<i>Exact</i>	5% <i>Type I</i>	<i>Normal</i>	<i>Exact</i>	1% <i>Type I</i>	<i>Normal</i>
5	.753	.742	.781	.798	.858	1.000
10	.561	.562	.572	.705	.702	.763
15	.462	.461	.466	.597	.596	.629
20	.399	.399	.401	.524	.524	.545
25	.356	.356	.357	.473	.473	.487
30	.324*	.324	.324	.433	.433	.444
45	.262	.262	.262	.356	.356	.362
75	.201*	.201	.201	.276	.276	.278

4. **Alternative expansions of the generating functions.** In this section we shall determine the exact generating functions which were approximated in (3.12) and (3.16) and obtain these same approximations in another manner. This development will enable us to see how good the approximation is in the sense that it gives a certain number of exact moments. The determinant in (3.4) for lag l and mean zero can be written

$$(4.1) \quad A_n = \begin{vmatrix} a & b & & & & & b \\ & b & a & b & & & \\ & & b & a & b & & 0 \\ & & & \ddots & \ddots & \ddots & \\ & & & & \ddots & \ddots & \\ & & 0 & & & b & a & b \\ b & & & & & & b & a \end{vmatrix}_n$$

where $a = 1 - 2t_2$, $b = -t_1$ and all the other elements are zero. The b in the upper right corner and lower left corner indicate the value b in the a_{1n} and a_{n1} positions. Let us define the following determinants:

$$(4.2) \quad B_n = \begin{vmatrix} a & b & & & & \\ b & a & b & & & \\ & b & a & b & & 0 \\ & & \ddots & \ddots & \ddots & \\ & & & 0 & & \\ & & & & b & a & b \\ b & & & & & b & a \end{vmatrix}_n, \quad C_n = \begin{vmatrix} b & a & b & & & \\ & b & a & b & & \\ & & b & a & b & 0 \\ & & & \ddots & \ddots & \\ & & & & 0 & \\ & & & & & b & a \end{vmatrix}_n$$

$$D_n = \begin{vmatrix} a & b & & & & \\ b & a & b & & & \\ & b & a & b & & 0 \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & \\ & & 0 & & & b & a & b \\ & & & & & b & a \end{vmatrix}_n$$

We see that

$$(4.3) \quad \begin{aligned} A_n &= B_n + (-1)^{n-1} b C_{n-1}, \\ B_n &= D_n + (-1)^{n-1} b^n, \\ C_n &= b^n + (-1)^{n-1} b D_{n-1}, \end{aligned}$$

and A_n can be expressed in terms of D_n by substituting for B_n and C_n in the equation for A_n .

$$(4.4) \quad A_n = D_n - b^2 D_{n-2} + 2(-1)^{n-1} b^n.$$

We can obtain an expression for D_n if we expand this determinant by the first row. This gives

$$(4.5) \quad D_n = a D_{n-1} - b^2 D_{n-2}.$$

Since this is a second order difference equation, the solution may be written $D_n = k_1 u^n + k_2 v^n$ where u, v are roots of the equation $x^2 - ax + b^2 = 0$. Now,

$$(4.6) \quad \begin{aligned} D_1 &= u + v = k_1 u + k_2 v, \\ D_2 &= u^2 + v^2 + uv = k_1 u^2 + k_2 v^2 \end{aligned}$$

so that we can determine k_1 and k_2 . We now see that

$$(4.7) \quad D_n = \frac{u^{n+1} - v^{n+1}}{u - v}$$

which upon substitution in the equation for A_n gives

$$(4.8) \quad \begin{aligned} A_n &= u^n + v^n + 2(-1)^{n-1}b^n, \\ &= u^n + v^n - 2t_1^n, \end{aligned}$$

where $u, v = \frac{1}{2}(1 - 2t_2) \pm \sqrt{(1 - 2t_2)^2 - 4t_1^2}$. Now $\phi_1(t_1, t_2) = A_n^{-1}$ and it is easily seen from the form of A_n directly above that derivatives up to the n th order with respect to t_1 in which t_1 is then set equal to zero will be given by derivatives of $A_n = u^n$ and this is the approximation (3.13) found by other methods.

The determinant in (3.4) for lag 1 and mean not equal to zero can be written

$$(4.9) \quad A_n = \begin{vmatrix} a & b & & & b \\ b & a & b & & c \\ & b & a & b & \\ & & \ddots & \ddots & \\ c & & & & b & a & b \\ b & & & & & b & a \end{vmatrix} \quad \begin{aligned} a &= 1 - 2t_2 + 2(t_1 + t_2)/n, \\ b &= -t_1 + 2(t_1 + t_2)/n, \\ c &= 2(t_1 + t_2)/n. \end{aligned}$$

Let us define the following determinants:

$$(4.10) \quad \begin{aligned} B_n &= \begin{vmatrix} b & a & b & & \\ & b & a & b & c \\ & & b & a & b \\ & & & \ddots & \ddots \\ & c & & & \\ & & & & b & a \\ b & & & & & b \end{vmatrix} & C_n &= \begin{vmatrix} a & b & & & \\ b & a & b & & c \\ & b & a & b & \\ & & \ddots & \ddots & \\ c & & & & \\ & & & & b & a & b \\ b & & & & & b & a \end{vmatrix} \\ D_n &= \begin{vmatrix} b & a & b & & \\ & b & a & b & c \\ & & b & a & b \\ & & & \ddots & \ddots \\ c & & & & \\ & & & & b & a \\ & & & & & b \end{vmatrix} & E_n &= \begin{vmatrix} a & b & & & \\ b & a & b & & c \\ & b & a & b & \\ & & \ddots & \ddots & \\ c & & & & \\ & & & & b & a & b \\ & & & & & b & a \end{vmatrix} \end{aligned}$$

If we replace the b in the upper right corner of A_n by $c + (b - c)e^t$ we obtain

$$(4.11) \quad A'_n = C'_n + (-1)^{n-1}(b - c)B_{n-1}.$$

If we replace the b in the lower left corner of B_n and C'_n by $c + (b - c)e^t$ we obtain

$$(4.12) \quad \begin{aligned} B_n &= D_n + (-1)^{n-1}(b - c)E_{n-1}, \\ C'_n &= E_n + (-1)^{n-1}(b - c)D_{n-1}. \end{aligned}$$

We now have A_n in terms of D_n and E_n . We must now evaluate D_n and E_n .

$$(4.13) \quad D_n = \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & b & a & b & \\ 0 & b & a & b & c \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c & & & & \\ & & & & b & a \\ & & & & b & a_{n+1} \\ 0 & & & & & & r & s \\ & & & & & & r & s_{n+1} \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ -c & r & s & r & \\ -c & r & s & r & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & & & & r & s \\ -c & & & & r & s_{n+1} \end{vmatrix}$$

where $r = b - c$ and $s = a - c$ and the second determinant above is obtained from the first by subtracting c times the first row from all other rows. Writing this last determinant as the sum of two determinants by separating the first column we get

$$(4.14) \quad D_n = r^n - cF_{n+1}.$$

Combining the above difference equations we obtain

$$(4.15) \quad A_n = E_n - r^2E_{n-2} + 2(-1)^nrcF_n + 2(-1)^{n-1}r^n$$

and see that we must obtain E_n and F_n .

Expansion of F_{n+1} by the second column gives

$$(4.16) \quad F_{n+1} = -G_n + rF_n$$

and expanding G_n by the last row we get

$$(4.17) \quad G_n = rG_{n-1} + (-1)^{n-1}H_{n-1}.$$

$$(4.18) \quad F_{n+1} = \begin{vmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & r & s & r & 0 \\ 1 & r & s & r & \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & & & & r & s \\ 1 & & & & r & s_{n+1} \end{vmatrix} \quad G_n = \begin{vmatrix} 1 & s & r \\ 1 & r & s & r & 0 \\ 1 & r & s & r & \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & & & & r & s \\ 1 & & & & r & s_n \end{vmatrix}$$

$$H_n = \begin{vmatrix} s & r \\ r & s & r & 0 \\ r & s & r & \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & & & r & s & r \\ & & & r & s \end{vmatrix}$$

H_n is the same type as (4.1), therefore $H_n = \frac{u^{n+1} - v^{n+1}}{u - v}$ where u and v are the roots of the equation $x^2 - sx + r^2 = 0$, so that (4.17) becomes

$$(4.19) \quad G_n - rG_{n-1} = (-1)^{n-1} \frac{u^n - v^n}{u - v}$$

and the solution of this equation gives

$$(4.20) \quad G_n = \frac{r^n}{2r + s} + \frac{(-1)^{n-1}[r(u^n - v^n) + u^{n+1} - v^{n+1}]}{(u - v)(2r + s)}.$$

Introducing this expression into (4.16) we find

$$(4.21) \quad F_n = (-1)^{n-1} \frac{u^n - v^n}{(u - v)(2r + s)} - \frac{nr^{n-1}}{2r + s}$$

$$(4.22) \quad E_n = \begin{vmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 0 & a & b & & & & \\ 0 & b & a & b & c & & \\ \cdot & \cdot & \cdot & & & & \\ \cdot & & \cdot & \cdot & \cdot & & \\ & & c & & & & \\ & & & b & a & b & \\ 0 & & & b & a & & \end{vmatrix}_{n+1} = \begin{vmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & 1 \\ -c & s & r & & & & \\ -c & r & s & r & 0 & & \\ \cdot & \cdot & \cdot & & & & \\ \cdot & & \cdot & \cdot & \cdot & & \\ & & 0 & & & & \\ & & & r & s & r & \\ -c & & & r & s & & \end{vmatrix}_{n+1}$$

where the second determinant is obtained from the first by subtracting c times the first row from all other rows. If we separate this last determinant on the first column we get

$$(4.23) \quad E_n = I_n - cI_{n+1}$$

$$(4.24) \quad I_n = \begin{vmatrix} 0 & 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 1 & s & r & & & & \\ 1 & r & s & r & 0 & & \\ \cdot & & \cdot & \cdot & & & \\ \cdot & & \cdot & \cdot & & & \\ & & 0 & & & & \\ & & & r & s & r & \\ 1 & & & r & s & & \end{vmatrix} \quad J_n = \begin{vmatrix} 0 & 1 & 1 & \cdot & \cdot & \cdot & 1 & 1 \\ 1 & s & r & & & & & \\ 1 & r & s & r & 0 & & & \\ \cdot & \cdot & \cdot & \cdot & & & & \\ \cdot & & \cdot & \cdot & & & & \\ & & 0 & & r & s & r & \\ 1 & & & r & s & & & \\ & & & & r & r & & \end{vmatrix}$$

Expanding I_n by the last row, we get

$$(4.25) \quad I_n = (-1)^{n-1}G_{n-1} - rJ_{n-1} + sI_{n-1}.$$

Expanding J_n by the last column, we get

$$(4.26) \quad J_n = (-1)^{n-1}G_{n-1} + rI_{n-1}.$$

If we combine these last two equations we find

$$(4.27) \quad I_n - sI_{n-1} + r^2I_{n-2} = (-1)^{n-1}(G_{n-1} + rG_{n-2}).$$

If we now solve this difference equation for I_n , substitute this solution in the equation for E_n , and in turn substitute this result and the expression we obtained for F_n in (4.15), we get

$$(4.28) \quad \begin{aligned} A_n &= \frac{u^n + v^n + 2(-1)^{n-1}r^n}{2r + s} \\ &= \frac{u^n + v^n - 2t_1^n}{1 - 2t_1 - 2t_2}. \end{aligned}$$

The final form results since $2r + s = 1 - 2t_1 - 2t_2$, $r = -t_1$. u and v have the same values as before. If we compare this result with that obtained in (4.8) for mean equal to zero we see that this is the same except that here we have the added factor, $1 - 2t_1 - 2t_2$, in the denominator. We have a similar result then for the approximation for derivatives of $\varphi_1(t_1, t_2) = A_n^{-\frac{1}{2}}$ for $t_1 = 0$. Here this approximation is $A_n = u^n/(1 - 2t_1 - 2t_2)$, the same result as that obtained in (3.18). This approximation will yield moments which are exact for $n > \alpha k$ for any lag l where α is the largest common factor of n and l . The reason for this restriction is easily seen if we consider the expansion obtained in (3.7), for

$$(4.29) \quad \frac{\partial \varphi(t_1, t_2)}{\partial t_1} = -\frac{1}{2} \varphi(t_1, t_2) \sum_{k=1}^n \frac{-2 \cos \frac{2\pi l k}{n}}{1 - 2t_2 - 2t_1 \cos \frac{2\pi l k}{n}}$$

with $t_1 = 0$,

$$(4.30) \quad \begin{aligned} \left. \frac{\partial \varphi(t_1, t_2)}{\partial t_1} \right]_{t_1=0} &= -\frac{1}{2} \varphi(0, t_2) \sum_{k=1}^n \frac{-2 \cos \frac{2\pi l k}{n}}{1 - 2t_2} \\ &= \frac{\varphi(0, t_2)}{1 - 2t_2} \sum_{k=1}^n \cos \frac{2\pi l k}{n}. \end{aligned}$$

Further

$$(4.31) \quad \left. \frac{\partial^2 \varphi(t_1, t_2)}{\partial t_1^2} \right]_{t_1=0} = \frac{\varphi(0, t_2)}{(1 - 2t_2)^2} \left[\left(\sum_{k=1}^n \cos \frac{2\pi l k}{n} \right)^2 + 2 \sum_{k=1}^n \cos^2 \frac{2\pi l k}{n} \right]$$

and the m th partial derivative will contain the sum of the m th powers of the cosines. These are the sums of the powers of the real parts of the roots of unity and it is easily seen that $\sum \cos^m \frac{2\pi l k}{n} = \sum \cos^m \frac{2\pi k}{n}$ only for $m < \alpha k$ where α is the largest common factor of n and l .

To change the moment generating function of \hat{b}_n to that of \hat{b} we must drop the last term of the product. In the above expressions we then have $\sum_{k=1}^{n-1} \cos^m \frac{2\pi l k}{n}$ and the same conclusion will hold.

5. Application to successive differences. If we change slightly the function $\eta = \delta_{n-1}^2/V_n$ investigated by von Neumann and Williams we find the moments and distribution greatly simplified. Let us define

$$(5.1) \quad \delta_n^2 = \sum_{i=1}^n (x_i - x_{i+1})^2$$

where $x_{n+1} = x_1$ and consider the ratio ${}_0\eta_1$ of δ_n^2 to Σx_i^2 . Now,

$$(5.2) \quad \delta_n^2 = 2\Sigma x_i^2 - 2\Sigma x_i x_{i+1}$$

therefore

$$(5.3) \quad {}_0\eta_1 = \frac{\delta_n^2}{\Sigma x_i^2} = 2(1 - \hat{b}_0)$$

and we may find the moments and distribution of ${}_0\eta_1$ directly from those of \hat{b}_0 . We find the moments to be:

$$(5.4) \quad \begin{aligned} m_1 &= 2 & m_2 &= \frac{2^2(n+3)}{n+2} \\ m_3 &= \frac{2^3(n+5)}{n+2} & m_4 &= \frac{2^4(n+5)(n+7)}{(n+2)(n+4)} \\ m_k &= \frac{2^k(n+2k-1)!}{(n+k)!(n+2)(n+4)\cdots(n+2k-2)}, & (k < n) \end{aligned}$$

and the ratio ${}_0\eta_1$ is distributed according to the law

$$(5.5) \quad C_{10}\eta_1^{1(n-1)}(4 - {}_0\eta_1)^{1(n-1)}$$

up to n moments

If we replace x_i in the above ratios by $x_i - \bar{x}$ we find the moments of the ratio $\eta_1 = \delta_n^2/\Sigma(x_i - \bar{x})^2$ to be:

$$(5.6) \quad \begin{aligned} m_1 &= \frac{2n}{n-1} & m_2 &= \frac{2^2 n(n+3)}{(n-1)(n+1)} \\ m_3 &= \frac{2^3 n(n+4)(n+5)}{(n-1)(n+1)(n+3)} & m_4 &= \frac{2^4 n(n+5)(n+6)(n+7)}{(n-1)(n+1)(n+3)(n+5)} \\ m_k &= \frac{2^k n(n+2k-1)!}{(n+k)!(n-1)(n+1)(n+3)\cdots(n+2k-3)} \end{aligned}$$

and $(\eta_1 - 2)^2 = z$ has the distribution

$$(5.7) \quad C_2 z^{-1/2} (4 - z)^{1/2(n-2)}$$

up to $\frac{1}{2}n$ moments.

The ratio $\delta_n^2/\Sigma x_i^2$ compares the variation of the first differences to that of the original variates. We might wish to compare the variation of the second

differences to that of the first differences. For this purpose let us form the ratio

$$(5.8) \quad \eta_2 = \frac{\sum_{i=1}^n (x_i - 2x_{i+1} + x_{i+2})^2}{\sum_{i=1}^n (x_i - x_{i+1})^2} \quad \begin{matrix} x_{n+1} = x_1 \\ x_{n+2} = x_2 \end{matrix}$$

to test the hypothesis H_2 that the variation of the second differences compared to the variation of the first differences is such as would occur by chance. Let x_1, x_2, \dots, x_n be normally distributed with mean μ and variance σ^2 . The ratio η_2 is independent of the mean value of the variates, therefore we may consider a distribution with mean equal zero. We shall develop the mean and variance of η_2 when the hypothesis to be tested is true. The moment generating function for the joint distribution of $D_2 = \Sigma(x_i - 2x_{i+1} + x_{i+2})^2/2\sigma^2$ and $D_1 = \Sigma(x_i - x_{i+1})^2/2\sigma^2$ is

$$(5.9) \quad \begin{aligned} \varphi(t_1, t_2) &= E[\exp(D_2 t_1 + D_1 t_2)] \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(D_2 t_1 + D_1 t_2 - \frac{1}{2\sigma^2} \Sigma x_i^2\right) \prod_{i=1}^n dx_i. \end{aligned}$$

We may find the moments then by a process similar to that outlined in (3.2) and (3.3). The next few steps are identical with (3.4) and (3.5). For the present problem, however, $a_1 = 1 - 6t_1 - 2t_2$, $a_2 = 4t_1 + t_2$, $a_3 = t_1$ so that

$$(5.10) \quad \begin{aligned} \varphi^{-2}(t_1, t_2) &= \prod_{k=1}^n [1 - 6t_1 - 2t_2 + (4t_1 + t_2)(\omega_k^4 + \omega_k^{-4}) + t_1(\omega_k^2 + \omega_k^{-2})], \\ &= \prod_{k=1}^n \left[1 - 6t_1 - 2t_2 + (8t_1 + 2t_2) \cos \frac{2\pi k}{n} + 2t_1 \cos \frac{4\pi k}{n}\right], \\ &= \prod_{k=1}^n \left[a + b \cos \frac{2\pi k}{n} + c \cos \frac{4\pi k}{n}\right]. \end{aligned}$$

If we follow the same procedure indicated in (3.8) to (3.13) we obtain successively

$$(5.11) \quad \varphi(t_1, t_2) = \prod_{k=1}^n (a + b \cos \alpha_k + c \cos 2\alpha_k)^{-1}$$

$$(5.12) \quad = e^{-\frac{1}{2} \sum_{k=1}^n \log(a + b \cos \alpha_k + c \cos 2\alpha_k)}$$

$$(5.13) \quad = \exp\left(-\frac{n}{4\pi} \sum_{k=1}^n \log(a + b \cos \alpha_k + c \cos 2\alpha_k)\right)$$

and replace the summation by the integral which is the limit of the summation as $n \rightarrow \infty$,

$$(5.14) \quad \varphi(t_1, t_2) \sim \exp\left(-\frac{n}{4\pi} \int_0^{2\pi} \log(a + b \cos \alpha + c \cos 2\alpha) d\alpha\right)$$

$$(5.15) \quad \sim \exp\left(-\frac{n}{2} \log\left[\kappa \frac{1 + \sqrt{1 - \delta^2}}{2} \frac{1 + \sqrt{1 - \eta^2}}{2}\right]\right)$$

where $\kappa = a - c; \quad \eta, \delta = \frac{b \pm \sqrt{b^2 + 8c^2 - 8ac}}{2(a - c)}.$

We then have approximately

$$(5.16) \quad \varphi(t_1, t_2) \sim [\tfrac{1}{4}\kappa(1 + \sqrt{1 - \delta^2})(1 + \sqrt{1 - \eta^2})]^{-1n}.$$

(5.14) follows from (5.13) if we replace the summation by an integral, and (5.15) is obtained in the following manner: replace $\cos 2\alpha$ by $2 \cos^2 \alpha - 1$ and factor the resulting quadratic and integrate the factors separately

$$\begin{aligned} \int_0^{2\pi} \log (a + b \cos \alpha + c \cos 2\alpha) d\alpha &= \int_0^{2\pi} \log (a - c + b \cos \alpha + 2c \cos^2 \alpha) d\alpha \\ (5.17) \quad &= \int_0^{2\pi} \log \kappa d\alpha + \int_0^{2\pi} \log (1 + \delta \cos \alpha) d\alpha + \int_0^{2\pi} \log (1 + \eta \cos \alpha) d\alpha \\ &= 2\pi \log \kappa + 2\pi \log \tfrac{1}{2}(1 + \sqrt{1 - \delta^2}) + 2\pi \log \tfrac{1}{2}(1 + \sqrt{1 - \eta^2}). \end{aligned}$$

If we now expand (5.16) by multiplying the factors within the brackets and substitute for κ , η and δ we find

$$(5.18) \quad \varphi(t_1, t_2) \sim [A + B + C + D]^{-1n} = P^{-1n},$$

where

$$\begin{aligned} a &= 1 - 6t_1 - 2t_2, & A &= \tfrac{1}{4}(1 - 4t_1 - 2t_2), \\ b &= 8t_1 + 2t_2, & B &= \tfrac{1}{4}[1 - 12t_1 - 4t_2 + 8t_1t_2 + 2t_2^2 \\ & & & - 2(4t_1 + t_2)\sqrt{t_2^2 + 4t_1}]^{\frac{1}{2}}, \\ (5.19) \quad c &= -2t_1, & C &= \tfrac{1}{4}[1 - 12t_1 - 4t_2 + 8t_1t_2 + 2t_2^2 \\ & & & + 2(4t_1 + t_2)\sqrt{t_2^2 + 4t_1}]^{\frac{1}{2}}, \\ & & D &= \tfrac{1}{4}(1 - 16t_1 - 4t_2)^{\frac{1}{2}}. \end{aligned}$$

From (5.18) $P = A + B + C + D$ and at $t_1 = 0$

$$\begin{aligned} P &= \tfrac{1}{4}(1 + \sqrt{1 - 4t_2})^2, \\ (5.20) \quad \frac{\partial P}{\partial t_1} &= -2[1 + 2(1 - 4t_2)^{-\frac{1}{2}}], \\ \frac{\partial^2 P}{\partial t_1^2} &= \frac{-32}{(1 - 4t_2)^{\frac{3}{2}}} - \frac{(1 - \sqrt{1 - 4t_2})^2}{2t_2^2}. \end{aligned}$$

Now

$$\begin{aligned} (5.21) \quad \frac{\partial \varphi}{\partial t_1} &= -\tfrac{1}{2}nP^{-1n-1} \frac{\partial P}{\partial t_1} \\ \frac{\partial^2 \varphi}{\partial t_1^2} &= -\tfrac{1}{2}nP^{-1n-2} \left[\left(-\tfrac{1}{2}n - 1\right) \left(\frac{\partial P}{\partial t_1}\right)^2 + P \frac{\partial^2 P}{\partial t_1^2} \right]. \end{aligned}$$

If we substitute in this formula and integrate the first with respect to t_2 we shall obtain the first moment of the ratio η_2 . If we integrate the second twice with respect to t_2 , we shall obtain the second moment of the ratio η_2 . We find these moments to be

$$(5.22) \quad \begin{aligned} M_1 &= \frac{3n+2}{n+1} & M_2 &= \frac{9n^2+23n+12}{(n+1)(n+2)} \\ \sigma^2 &= \frac{2n^2+7n+4}{(n+1)^2(n+2)}. \end{aligned}$$

6. Likelihood criteria for multiple serial correlation. Given a sample of n observations, x_1, x_2, \dots, x_n , we shall assume that they are distributed according to the law

$$(6.1) \quad dP_n = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{a=1}^n \left(x_a - a - \sum_{i=1}^r b_i x_{a-i} \right)^2} dx_1 \cdots dx_n \quad (x_i = x_{-i+1})$$

that is, that each variate, say at the time t , has as its mean value a linear function of the variates at time $t-l_1, t-l_2$, etc. Let us investigate the likelihood criteria for testing the hypothesis, H_r , that each variate is independent of the others; i.e. that the $b_i = 0$ ($i = 1, \dots, r$). For the hypothesis H_r we define the space Ω and the space ω , as follows:

$$(6.2) \quad \begin{cases} \Omega: & \sigma^2 > 0 & -\infty < a, b_i < \infty \\ \omega: & \sigma^2 > 0 & -\infty < a < \infty, \quad b_i = 0, \end{cases}$$

we find the likelihood ratio criterion

$$(6.3) \quad \lambda_r^{2/n} = \frac{|a_{ij}|}{a_{00}|a_{pq}|} \quad \begin{matrix} i, j = 0, 1, \dots, r \\ p, q = 1, \dots, r \end{matrix}$$

in which

$$(6.4) \quad \begin{aligned} a_{00} &= \sum_{\alpha} (x_{\alpha} - \bar{x})^2 \\ a_{0i} &= \sum_{\alpha} (x_{\alpha} - \bar{x})(x_{\alpha+l_i} - \bar{x}) \\ a_{ij} &= \sum_{\alpha} (x_{\alpha+l_i} - \bar{x})(x_{\alpha+l_j} - \bar{x}) \end{aligned}$$

and it is noted that $a_{ii} = a_{00}$ and if the l_i are equispaced $a_{i,j} = a_{00}$. λ_r is a statistic which measures how completely each variate at time t can be expressed as a linear function of variates spaced at time $t-l_1, t-l_2$, etc.

Next we shall develop a statistic for testing the hypothesis, $H_{r,m}$, that of the set of the values b_i ($i = 1, \dots, r$) in (6.1) the subset $b_{m+1}, b_{m+2}, \dots, b_r = 0$. Here we have the same likelihood function but for $H_{r,m}$, we define the spaces Ω and ω as follows:

$$(6.5) \quad \begin{cases} \Omega: & \sigma^2 > 0 & -\infty < a, b_i < \infty; \\ \omega: & \sigma^2 > 0 & -\infty < a, b_u < \infty, \quad b_w = 0, \end{cases}$$

$$u = (1, \dots, m), \quad w = (m+1, \dots, r),$$

and obtain the criterion

$$(6.6) \quad \lambda_{r,m}^{2/n} = \frac{|a_{ij}| |a_{uv}|}{|a_{pq}| |a_{st}|}$$

$$\begin{aligned} i, j &= 0, 1, \dots, r, \\ p, q &= 1, \dots, r, \\ s, t &= 0, 1, \dots, m, \\ u, v &= 1, \dots, m, \\ &(m < r). \end{aligned}$$

The form and the derivation of these λ criteria parallels very closely that of the likelihood ratio criteria obtained in multivariate analysis for testing significance of regression coefficients.

CASE I. If we set $r = 1$ in λ_r we obtain

$$(6.7) \quad \lambda_1^{2/n} = \frac{\begin{vmatrix} a_{00} & a_{01} \\ a_{01} & a_{00} \end{vmatrix}}{a_{00} \ a_{00}} = 1 - \frac{a_{01}^2}{a_{00}^2} = 1 - b^2,$$

for which the distribution is given in (3.24).

CASE II. If we set $r = 2$, we have

$$(6.8) \quad \lambda_2^{2/n} = \frac{\begin{vmatrix} a_{00} & a_{01} & a_{02} \\ a_{01} & a_{11} & a_{12} \\ a_{02} & a_{12} & a_{22} \end{vmatrix}}{a_{00} \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}},$$

for which if we take $l_1 = 1$, $l_2 = 2$ we get

$$(6.9) \quad \frac{\begin{vmatrix} a_{00} & a_{01} & a_{02} \\ a_{01} & a_{00} & a_{01} \\ a_{02} & a_{01} & a_{00} \end{vmatrix}}{a_{00} \begin{vmatrix} a_{00} & a_{01} \\ a_{01} & a_{00} \end{vmatrix}}.$$

The expanded form of this numerator is $a_{00}^3 + 2a_{01}^2a_{02} - a_{00}a_{02}^2 - 2a_{01}^2a_{00}$.

Let us consider

$$(6.10) \quad \varphi(\theta_0, \theta_1, \theta_2) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(\sum x_a^2(1-\theta_0) - \theta_1 \sum x_a x_{a+1} - \theta_2 \sum x_a x_{a+2})} \Pi dx_a.$$

We shall find the mean and variance of ${}_0\lambda_2^{2/n}$ (mean = 0) when the hypothesis ${}_0H_r$ ($r = 2$) is true. We can find the first moment of ${}_0\lambda_2^{2/n}$ then by performing the following operations: (a) compute

$$(6.11) \quad \frac{\partial^3 \varphi}{\partial \theta_0^3} + 2 \frac{\partial^2 \varphi}{\partial \theta_1^2} \frac{\partial \varphi}{\partial \theta_2} - \frac{\partial \varphi}{\partial \theta_0} \frac{\partial^2 \varphi}{\partial \theta_2^2} - 2 \frac{\partial^2 \varphi}{\partial \theta_1^2} \frac{\partial \varphi}{\partial \theta_0},$$

(this will give the first moment of the numerator) and set $\theta_2 = 0$, (b) integrate from $-\infty$ to $\theta_0 + \theta_1$ with respect to $\theta_0 + \theta_1 = \xi$, from $-\infty$ to $\theta_0 - \theta_1$, with respect to $\theta_0 - \theta_1 = \xi$, and set $\theta_1 = 0$ (at this point we will have the first moment of the third order determinant divided by the second order determinant), (c)

integrate with respect to θ_0 from $-\pi$ to 0. The reason for step (b) is easily seen since the second order determinant $a_{ij}^2 = a_{ij}^2$ may be written $(a_{i1} + a_{i0})$ ($a_{00} + a_{01}$).

Further moments may be computed in a similar manner. $a_{ij} = a_{ij}(\theta_0, \theta_1, \theta_2)$ may be written as a determinant in the manner indicated in (3.4) and (3.5). Here, $a_1 = 1 - \theta_0$, $a_2 = a_n = -\frac{1}{2}\theta_1$ and $a_3 = a_{n-1} = -\frac{1}{2}\theta_2$ and a_0 to $a_{n-2} = 0$, then

$$(6.12) \quad {}_0\varphi_2^{-2}(\theta_0, \theta_1, \theta_2) = \prod_{k=1}^n \sum_{i=1}^n a_i \omega_k^{i-1} = \prod_{k=1}^n \left(a_1 + 2a_2 \cos \frac{2\pi k}{n} + 2a_3 \cos \frac{4\pi k}{n} \right).$$

We shall approximate ${}_0\varphi_2(\theta_0, \theta_1, \theta_2)$ by the method contained between (5.11) and (5.18). We set

$$(6.13) \quad {}_0\varphi_2(\theta_0, \theta_1, \theta_2) = \prod_{k=1}^n \left(a + b \cos \frac{2\pi k}{n} + c \cos \frac{4\pi k}{n} \right)^{-1}$$

and obtained

$$(6.14) \quad {}_0\varphi_2(\theta_0, \theta_1, \theta_2) \sim [A + B + C + D]^{-1/n} P^{-1/n}$$

where

$$(6.15) \quad \begin{aligned} A &= \frac{1}{4}(a - c) \\ B &= \frac{1}{8}(4a^2 - 4c^2 - 2b^2 - 2bE)^{\frac{1}{2}} = \frac{1}{8}b^{\frac{1}{2}} \\ C &= \frac{1}{8}(4a^2 - 4c^2 - 2b^2 + 2bE)^{\frac{1}{2}} = \frac{1}{8}\gamma^{\frac{1}{2}} \\ D &= \frac{1}{4}((a + c)^2 - b^2)^{\frac{1}{2}} = \frac{1}{4}\epsilon^{\frac{1}{2}} \\ E &= b^2 + 8c(c - a). \end{aligned}$$

It is easily seen that we may operate (differentiate and integrate) with a, b, c , in place of $\theta_0, \theta_1, \theta_2$ respectively. Therefore, we compute

$$(6.16) \quad \begin{aligned} \frac{\partial \varphi}{\partial c} &= -\frac{1}{2}nP^{-1/n-1} \frac{\partial P}{\partial c} \\ \frac{\partial^2 \varphi}{\partial c^2} &= -\frac{1}{2}nP^{-1/n-2} \left[(-\frac{1}{2}n - 1) \left(\frac{\partial P}{\partial c} \right)^2 + P \frac{\partial^2 P}{\partial c^2} \right] \end{aligned}$$

and since $P = A + B + C + D$ we compute

$$(6.17) \quad \begin{aligned} \frac{\partial A}{\partial c} &= -\frac{1}{4}; & \frac{\partial^2 A}{\partial c^2} &= 0 \\ \frac{\partial B}{\partial c} &= \frac{1}{16}\beta^{-1} \frac{\partial \beta}{\partial c}; & \frac{\partial^2 B}{\partial c^2} &= \frac{1}{16} \left[-\frac{1}{2}\beta^{-1} \left(\frac{\partial \beta}{\partial c} \right)^2 + \beta^{-1} \frac{\partial^2 \beta}{\partial c^2} \right] \\ \frac{\partial C}{\partial c} &= \frac{1}{16}\gamma^{-1} \frac{\partial \gamma}{\partial c}; & \frac{\partial^2 C}{\partial c^2} &= \frac{1}{16} \left[-\frac{1}{2}\gamma^{-1} \left(\frac{\partial \gamma}{\partial c} \right)^2 + \gamma^{-1} \frac{\partial^2 \gamma}{\partial c^2} \right] \\ \frac{\partial D}{\partial c} &= \frac{1}{8}\epsilon^{-1} \frac{\partial \epsilon}{\partial c}; & \frac{\partial^2 D}{\partial c^2} &= \frac{1}{8} \left[-\frac{1}{2}\epsilon^{-1} \left(\frac{\partial \epsilon}{\partial c} \right)^2 + \epsilon^{-1} \frac{\partial^2 \epsilon}{\partial c^2} \right] \\ \frac{\partial E}{\partial c} &= 16c - 8a; & \frac{\partial^2 E}{\partial c^2} &= 16. \end{aligned}$$

In order to evaluate the expressions in (6.17) we must find

$$\begin{aligned}
 \frac{\partial \beta}{\partial c} &= -8c - bE^{-1} \frac{\partial E}{\partial c}; & \frac{\partial^2 \beta}{\partial c^2} &= -8 - bE^{-1} \left[-\frac{1}{2} \left(\frac{\partial E}{\partial c} \right)^2 + E \frac{\partial^2 E}{\partial c^2} \right] \\
 (6.18) \quad \frac{\partial \gamma}{\partial c} &= -8c + bE^{-1} \frac{\partial E}{\partial c}; & \frac{\partial^2 \gamma}{\partial c^2} &= -8 + bE^{-1} \left[-\frac{1}{2} \left(\frac{\partial E}{\partial c} \right)^2 + E \frac{\partial^2 E}{\partial c^2} \right] \\
 \frac{\partial \epsilon}{\partial c} &= 2(a + c); & \frac{\partial^2 \epsilon}{\partial c^2} &= 2.
 \end{aligned}$$

If we now set $c = 0$, we obtain

$$\begin{aligned}
 P &= \frac{1}{2}(a + (a^2 - b^2)^{\frac{1}{2}}) \\
 (6.19) \quad \frac{\partial P}{\partial c} &= \frac{a - (a^2 - b^2)^{\frac{1}{2}}}{2(a^2 - b^2)^{\frac{1}{2}}} \\
 \frac{\partial^2 P}{\partial c^2} &= \frac{2a^4 - 4a^2b^2 + b^4 + (-2a^3 + 2ab^2)(a^2 - b^2)^{\frac{1}{2}}}{2b^2(a^2 - b^2)^{\frac{3}{2}}}.
 \end{aligned}$$

We may now substitute these values in (6.16) and then substitute the resulting expressions in (6.11). The remaining values that are required for (6.11) are easily computed since they may be obtained from φ with $c = 0$, i.e.

$$(6.20) \quad \varphi_2(\theta_0, \theta_1, 0) = [\frac{1}{2}(a + (a^2 - b^2)^{\frac{1}{2}})]^{-1n}.$$

The result of these substitutions gives

$$(6.21) \quad \frac{-n^2(n+3)P^{-\frac{1}{2}n-2}}{8(a^2 - b^2)^{\frac{1}{2}}},$$

in which we set $d = \frac{1}{2}(a - b)$ and $e = \frac{1}{2}(a + b)$ and integrate with respect to d and e . We obtain

$$(6.22) \quad \frac{-n^2}{2(n+2)} [\frac{1}{2}(a + (a^2 - b^2)^{\frac{1}{2}})]^{-\frac{1}{2}n-1},$$

and if we set $b = 0$ and integrate with respect to a , setting $a = 1$, ($\theta_0 = 0$), we finally have

$$(6.23) \quad E(\lambda_2^{2/n}) = \frac{n}{n+2}.$$

We shall now obtain the first moment of λ_2 without the restriction that the mean equal zero. For this purpose let us consider

$$(6.24) \quad \varphi_2(\theta_0, \theta_1, \theta_2) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-(1/2\sigma^2)[a_0\theta_0(1-\theta_0)-a_0\theta_1-a_0\theta_2]} \Pi \, dx_a.$$

Here $a_1 = 1 - \theta_0 + m$, $a_2 = a_n = -\frac{1}{2}\theta_1 + m$, $a_3 = a_{n-1} = -\frac{1}{2}\theta_2 + m$, a_4 to $a_{n-2} = m$ where $m = (\theta_0 + \theta_1 + \theta_2)/n$. Expanding the determinant as in (6.12) we find

$$\begin{aligned}
 (6.25) \quad \varphi_1^{-2}(\theta_0, \theta_1, \theta_2) &= \prod_{k=1}^n \sum_{i=1}^n a_i \omega_k^{i-1} \\
 &= \prod_{k=1}^n \left(a_1 + 2a_2 \cos \frac{2\pi k}{n} + 2a_3 \cos \frac{4\pi k}{n} + \sum_{i=4}^{n-2} a_i \omega_k^{i-1} \right).
 \end{aligned}$$

Now

$$(6.26) \quad \sum_{k=1}^{n-2} a_k \omega_k^{i-1} = m \sum_{k=1}^{n-2} \omega_k^{i-1} \\ = \begin{cases} -m(1 + \omega^i + \omega^{i-1} + \omega^2 + \omega^{-2}), & k \neq n, \\ m(n-5), & k = n, \end{cases}$$

so that

$$(6.27) \quad \varphi_x^{-2}(\theta_0, \theta_1, \theta_2) = \prod_{k=1}^{n-1} \left(a_k + 2a_2 \cos \frac{2\pi k}{n} + 2a_3 \cos \frac{4\pi k}{n} \right).$$

We have obtained here a product which is the same as that in (6.12) except that the last factor is missing. The approximation corresponding to (6.14) will then be

$$(6.28) \quad \varphi_2(\theta_0, \theta_1, \theta_2) \cong \frac{[A + B + C + D]^{-1/n}}{(a + b + c)^{-1}},$$

since we may take the approximation for the product from 1 to n and divide by the last factor, $(a + b + c)$. The procedure for finding the first moment for λ_2 (mean = a) is exactly the same as that outlined for finding the first moment of $\phi\lambda_2$ (mean = 0). We obtain

$$(6.29) \quad E(\lambda_2^{2/n}) = \frac{n-1}{n+1}.$$

CASE III. If we set $r = 2$, $m = 1$ in $\lambda_{r,m}$ we have, if we take $l_1 = 1$, $l_2 = 2$

$$(6.30) \quad \lambda_{2,1}^{2/n} = \frac{\begin{vmatrix} a_{00} & a_{01} & a_{02} \\ a_{01} & a_{00} & a_{01} \\ a_{02} & a_{01} & a_{00} \end{vmatrix} a_{00}}{\begin{vmatrix} a_{00} & a_{01} \\ a_{01} & a_{00} \end{vmatrix}^2} \dots$$

To find the moments of $\lambda_{2,1}$ let us consider the following distribution,

$$(6.31) \quad dP_n = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{\alpha=1}^n (x_\alpha - \bar{x} - \beta(x_{\alpha+1} - \bar{x}))^2} \Pi dx_\alpha,$$

in which β represents the population value of the serial correlation coefficient. The moment generating function for the joint distribution of $a_{00}/2\sigma^2$, $a_{01}/2\sigma^2$ and $a_{02}/2\sigma^2$ will be

$$(6.32) \quad \varphi_{2,1}(\theta_0, \theta_1, \theta_2) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \int_{-\infty}^{\infty} \dots \int \exp \left(-\frac{1}{2\sigma^2} \{ \Sigma [(x_\alpha - \bar{x}) - \beta(x_{\alpha+1} - \bar{x})]^2 \right. \\ \left. - a_{00}\theta_0 - a_{01}\theta_1 - a_{02}\theta_2 \} \right) \Pi dx_\alpha \\ = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \int_{-\infty}^{\infty} \dots \int \exp \left(-\frac{1}{2\sigma^2} [a_{00}(1 + \beta^2 - \theta_0) \right. \\ \left. + a_{01}(-2\beta - \theta_1) + a_{02}(-\theta_2)] \right) \Pi dx_\alpha.$$

This function is very similar to (6.24). The approximation to $\varphi_{2,1}(\theta_0, \theta_1, \theta_2)$ here will be exactly the same as that obtained in (6.28) for $\varphi_2(\theta_0, \theta_1, \theta_2)$ except that here $a = 1 + \beta^2 - \theta_0$, $b = -2\beta - \theta_1$, $c = -\theta_2$. For the case where the mean is zero, we find the approximation (6.14) in which a , b , and c have the above values.

We may obtain the first moment of $\lambda_{2,1}$ by operating on the function $\varphi_{2,1}(\theta_0, \theta_1, \theta_2)$ proceeding as follows: (a) compute (6.11) as before and set $\theta_2 = 0$, (b) integrate from $-\infty$ to $\theta_0 + \theta_1$ with respect to $\theta_0 + \theta_1 = \xi$ from $-\infty$ to $\theta_0 - \theta_1$ with respect to $\theta_0 - \theta_1 = \eta$ (at this point we will have the first moment of the third order determinant divided by the second order determinant), (c) differentiate with respect to θ_0 , (d) repeat step (b), and set θ_0 and $\theta_1 = 0$.

The first two steps for obtaining the first moment of ${}_0\lambda_{2,1}$ (mean = 0) were performed for the first moment of ${}_0\lambda_2$ so that we may perform step (c) on (6.22). We obtain

$$(6.33) \quad \frac{n^2[\frac{1}{2}(a + (a^2 - b^2)^{\frac{1}{2}})]^{-\frac{1}{2}n-1}}{4(a^2 - b^2)^{\frac{1}{2}}},$$

and finally by step (d) we have

$$(6.34) \quad E({}_0\lambda_{2,1}^{2/n}) = \frac{n}{n+1} [\frac{1}{2}(a + (a^2 - b^2)^{\frac{1}{2}})]^{-\frac{1}{2}n-1},$$

in which $a = 1 + \beta^2$ and $b = -2\beta$ since θ_0 and θ_1 have been set equal to zero. Substitution of these values in (6.34) shows that it is independent of β , and we find

$$(6.35) \quad E({}_0\lambda_{2,1}^{2/n}) = \frac{n}{n+1}.$$

Using $\varphi_{2,1}(\theta_0, \theta_1, \theta_2)$, the generating function for $\lambda_{2,1}$ (mean = a), we find

$$(6.36) \quad E(\lambda_{2,1}^{2/n}) = \frac{n-1}{n}.$$

The procedure for obtaining the second moments of the above criteria consists essentially of performing twice the operations prescribed for obtaining the first moments. The details given in connection with the first moment are sufficient to indicate the procedure. The details for the second moments are too complicated algebraically to list here. Table I indicates the second moments obtained as well as other moments obtained in the earlier parts of the paper.

7. Serial correlation in several variables. Given a sample of n observations on each of k variables $x_{i\alpha}$, $i = 1, \dots, k$, we shall assume they are distributed as follows:

$$(7.1) \quad dP_n = \frac{A^{\frac{1}{2}n}}{(2\pi)^{\frac{1}{2}nk}} e^{-\frac{1}{2} \sum_{i,j} A_{ij}(x_{i\alpha}-a_i-b_i x_{i,a+k})(x_{j\alpha}-a_j-b_j x_{j,a+k})} \prod dx_{i\alpha}.$$

We wish to test the hypothesis H_{kn} that there is no serial correlation, i.e., that $b_i = 0$, $i = 1, \dots, k$. For this purpose let us define the space Ω and ω as follows:

$$(7.2) \quad \begin{cases} \Omega: & ||A_{ij}|| \text{ pos. def. } -\infty < a_i, b_i < \infty \\ \omega: & ||A_{ij}|| \text{ pos. def. } -\infty < a_i < \infty; b_i = 0. \end{cases}$$

TABLE I

x	formula no	$E(x)$	$E(x^2)$	σ^2
\hat{b}_0	(2.11)	0	$\frac{1}{n+2}$	$\frac{1}{n+2}$
\hat{b}	(2.6)	$\frac{-1}{n-1}$	$\frac{1}{n+1}$	$\frac{n(n-3)}{(n-1)^2(n+1)}$
${}_{00}\eta_1$	(5.3)	2	$\frac{4(n+3)}{n+2}$	$\frac{4}{n+2}$
η_1	(5.6)	$\frac{2n}{n-1}$	$\frac{4n(n+3)}{(n-1)(n+1)}$	$\frac{4n(n-3)}{(n-1)^2(n+1)}$
η_2	(5.8)	$\frac{3n+2}{n+1}$	$\frac{9n^2+23n+12}{(n+1)(n+2)}$	$\frac{2n^2+7n+4}{(n+1)^2(n+2)}$
${}_{00}\lambda_1$	(2.10)ff.	$\frac{n+1}{n+2}$	$\frac{(n+1)(n+3)}{(n+2)(n+4)}$	$\frac{2(n+1)}{(n+2)^2(n+4)}$
λ_1	(2.10)	$\frac{n}{n+1}$	$\frac{n(n+2)}{(n+1)(n+3)}$	$\frac{2n}{(n+1)^2(n+3)}$
${}_{00}\lambda_2$	(6.9)ff.	$\frac{n}{n+2}$	$\frac{n}{n+4}$	$\frac{4n}{(n+2)^2(n+4)}$
λ_2	(6.9)	$\frac{n-1}{n+1}$	$\frac{n-1}{n+3}$	$\frac{4(n-1)}{(n+1)^2(n+3)}$
${}_{00}\lambda_{2,1}$	(6.30)ff.	$\frac{n}{n+1}$	$\frac{n(n+2)}{(n+1)(n+3)}$	$\frac{2n}{(n+1)^2(n+3)}$
$\lambda_{2,1}$	(6.30)	$\frac{n-1}{n}$	$\frac{(n-1)(n+1)}{n(n+2)}$	$\frac{2(n-1)}{n^2(n+2)}$

The mean of \hat{b}_0 and \hat{b} were also obtained by Anderson [8].

The development of the appropriate λ criterion for this case parallels very closely the development of the λ criteria in multiple regression analysis. The criterion obtained for testing the hypothesis H_{kn} is

$$(7.3) \quad \lambda_{kn}^{2/n} = \frac{\begin{vmatrix} a_{ij} & b_{ij} \\ b_{ij} & a_{ij} \end{vmatrix}}{|a_{ij}|^2},$$

where

$$a_{ij} = \sum_{\alpha} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j),$$

$$b_{ij} = \frac{1}{2}[\sum (x_{i\alpha} - \bar{x}_i)(x_{j,\alpha+1} - \bar{x}_j) + \sum (x_{i,\alpha+1} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)].$$

The probability theory for the λ criteria in (7.3) remains to be developed.

8. Summary. A problem in serial correlation which has received considerable attention is that of devising a statistic for indicating the presence of a relation between successive observations, i.e. a lack of independence of the order in which the observations were drawn. Von Neumann developed the distribution and moments of the ratio of the mean square successive difference to the variance. R. L. Anderson presented the distribution of a serial correlation coefficient which is the ratio $R = \Sigma x_\alpha x_{\alpha+l} / \Sigma x_\alpha^2$ ($l \geq 1$, subscripts reduced mod n).

The present investigation was undertaken with the object of developing the likelihood ratio functions for testing various hypotheses connected with serial correlation in one or more variables and determining the moments and in some cases the distributions of these likelihood ratios.

The variates are considered to be ordered by their subscripts $\alpha = 1, \dots, n$. The introduction of $x_{n+1} = x_1$, $x_{n+2} = x_2$ etc. is made to obtain a symmetry which greatly simplifies the problem.

The likelihood ratio criteria were developed for testing the hypotheses

- a) that x_α is independent of $x_{\alpha+l}$
- b) that x_α is independent of $x_{\alpha+l_i}$, $i = 1, \dots, r$,
- c) that x_α is independent of some subset of the $x_{\alpha+l_i}$,
- d) that in the case of several variables $x_{i\alpha}$, $i = 1, \dots, k$, $\alpha = 1, \dots, n$ the $x_{i\alpha}$, $i = 1, \dots, k$ are independent of the $x_{i, \alpha+l_i}$. These criteria are similar in form to those obtained in regression analysis.

The likelihood ratio criterion for testing the hypothesis a) turns out to be $\lambda = (1 - R^2)^{n/2}$ where R is the function given above. The moments of R are obtained and from these the moments of $\lambda^{2/n}$. These moments are found to agree with those of a Pearson Type I curve to $n/2$ moments. A simple transformation gives us the moments of a ratio differing from that used by von Neumann by the addition of the term $(x_n - x_1)^2$ to the numerator. A simplification of the moments is attained by this change. In fact, if we denote this altered statistic by η we find that $(\eta - 2)^2$ is distributed according to a Pearson Type I curve to $n/2$ moments.

The mean and variance were determined for the ratio of the sum of squares of the second successive differences to the first successive differences.

The mean and variance are obtained for the likelihood criteria for testing the hypothesis b) for $r = 2$, and for testing the hypothesis c) for $r = 2$ where $x_{\alpha+l_2}$ is the subset of $x_{\alpha+l_i}$; ($i = 1, 2$).

All the above moments were obtained under the assumption that the hypothesis to be tested was true. No results have been obtained thus far in cases b) and c) for a general r nor for hypothesis d).

The moments for the several cases above were obtained by the use of moment generating functions which, for the criteria used, took the form of the product of n terms. In the case a) it was shown that the product could be approximately represented by the n th power of a single expression which was equivalent for the purpose of obtaining the first n moments. A method was developed for making analogous approximations to the generating functions for cases b) and c) since

it was not found possible to obtain the moments from the products in their original form.

The author wishes to express his gratitude to Professor S. S. Wilks under whose helpful direction this paper was written.

REFERENCES

- [1] J. VON NEUMANN, R. H. KENT, H. R. BELMONT, and B. I. HART, "The mean square successive difference," *Annals of Math. Stat.*, Vol. 12(1941) pp. 151-162.
- [2] J. VON NEUMANN, "Distribution of the ratio of the mean square difference to the variance," *Annals of Math. Stat.*, Vol. 12(1941) pp. 367-395.
- [3] B. I. HART and J. VON NEUMANN, "Tabulation of the probabilities of the ratio of the mean square successive difference to the variance," *Annals of Math. Stat.*, Vol. 13(1942) pp. 207-214.
- [4] J. VON NEUMANN, "A further remark concerning the distribution of the ratio of the mean square successive difference to the variance," *Annals of Math. Stat.*, Vol. 13(1942) pp. 86-88.
- [5] J. D. WILLIAMS, "Moments of the ratio of the mean square successive difference to the mean square difference in samples from a normal universe," *Annals of Math. Stat.*, Vol. 12(1941) pp. 239-241.
- [6] L. C. YOUNG, "On randomness in ordered sequences," *Annals of Math. Stat.*, Vol. 12(1941) pp. 293-300.
- [7] R. L. ANDERSON, "Serial correlation in the analysis of time series," unpublished thesis, Iowa State College, 1941.
- [8] R. L. ANDERSON, "Distribution of the serial correlation coefficient," *Annals of Math. Stat.*, Vol. 13(1942) pp. 1-13.
- [9] T. KOOPMANS, "Serial correlation and quadratic forms in normal variables," *Annals of Math. Stat.*, Vol. 13(1942) pp. 14-33.
- [10] P. S. LAPLACE, *Traité de Mécanique Céleste*, Vol. 2, paragraph 21.

ON A STATISTICAL PROBLEM ARISING IN THE CLASSIFICATION OF AN INDIVIDUAL INTO ONE OF TWO GROUPS¹

BY ABRAHAM WALD

Columbia University

1. Introduction. In social, economic and industrial problems we are often confronted with the task of classifying an individual into one of two groups on the basis of a number of test scores. For example, in the case of personnel selection the acceptance or rejection of an applicant is frequently based on a number of test scores obtained by the applicant. A similar situation arises in connection with college entrance examinations. Again, on the basis of a number of test scores, the admission or rejection of a student has to be decided. In all such problems it is assumed that there are two populations, say π_1 and π_2 , one representing the population of individuals fit, and the other the population of individuals unfit for the purpose under consideration. The problem is that of classifying an individual into one of the populations π_1 and π_2 on the basis of his test scores. Often, some statistical data from past experience are available which can be utilized in making the classification. Suppose that from past experience we have the test scores of N_1 individuals who are known to belong to population π_1 , and also the test scores of N_2 individuals who are known to belong to population π_2 . These data will be utilized in classifying a new individual on the basis of his test scores.

In this paper we shall deal with the statistical problem of classifying an individual into one of the populations π_1 and π_2 on the basis of his test scores and on the basis of past experience, given in the form of two samples, one drawn from π_1 and the other from π_2 . In the next section we give a precise formulation of the statistical problem and state the assumptions we make about the populations π_1 and π_2 .

2. Statement of the problem. We consider two sets of p variates (x_1, \dots, x_p) and (y_1, \dots, y_p) . It is assumed that each of the sets (x_1, \dots, x_p) and (y_1, \dots, y_p) has a p -variate normal distribution and the two sets are independent of each other. It is furthermore assumed that the covariance matrix of the variates x_1, \dots, x_p is equal to the covariance matrix of the variates y_1, \dots, y_p , i.e. $\sigma_{x_i x_j} = \sigma_{y_i y_j}$ ($i, j = 1, \dots, p$). We will denote this common covariance by σ_{ij} . Let us denote the mean value of x_i by μ_i and the mean value of y_i by ν_i . Furthermore we will denote the normal population with mean values μ_1, \dots, μ_p and covariance matrix $\|\sigma_{ij}\|$ by π_1 , and the normal population with mean values ν_1, \dots, ν_p and covariance matrix $\|\sigma_{ij}\|$ by π_2 .

A sample of size N_1 is drawn from the population π_1 and a sample of size N_2 is

¹The author wishes to thank Dr Irving Lorge, Columbia University, for calling his attention to this problem

drawn from the population π_1 . Denote by $x_{i\alpha}$ the α -th observation on x_i ($i = 1, \dots, p$; $\alpha = 1, \dots, N_1$) and $y_{i\beta}$ the β -th observation on y_i ($i = 1, \dots, p$; $\beta = 1, \dots, N_2$). Let z_i ($i = 1, \dots, p$) be a single observation on the i -th variate drawn from a p -variate population π , where it is known a priori that π is either identical with π_1 or with π_2 . The set (z_1, \dots, z_p) is assumed to be distributed independently of (x_1, \dots, x_p) and (y_1, \dots, y_p) .

We will deal here with the following statistical problem: On the basis of the observations $x_{i\alpha}, y_{i\beta}, z_i$ ($i = 1, \dots, p$; $\alpha = 1, \dots, N_1$; $\beta = 1, \dots, N_2$) we test the hypothesis H_1 that the population π , from which the set (z_1, \dots, z_p) has been drawn, is equal to π_1 . The parameters $\mu_1, \dots, \mu_p, \nu_1, \dots, \nu_p$ and $\|\sigma_{ij}\|$ are assumed to be unknown.

3. The statistic to be used for testing the hypothesis H_1 . In this problem there exists only a single alternative hypothesis to the H -hypothesis H_1 to be tested, i.e. the hypothesis H_2 that π is equal to π_2 . If the parameters $\mu_1, \dots, \mu_p, \nu_1, \dots, \nu_p$ and $\|\sigma_{ij}\|$ were known we could easily find (on the basis of a lemma by Neyman and Pearson) the critical region which is most powerful with respect to the alternative H_2 . Let us assume for the moment that the parameters $\mu_1, \dots, \mu_p, \nu_1, \dots, \nu_p$ and $\|\sigma_{ij}\|$ are known and let us compute the critical region for testing H_1 which is most powerful with respect to the alternative H_2 . According to a lemma by Neyman and Pearson² this critical region is given by the inequality

$$(1) \quad \frac{p_2(z_1, \dots, z_p)}{p_1(z_1, \dots, z_p)} \geq k,$$

where $p_1(z_1, \dots, z_p)$ denotes the joint probability density function of z_1, \dots, z_p under the hypothesis H_1 , $p_2(z_1, \dots, z_p)$ denotes the joint probability density function of (z_1, \dots, z_p) under the hypothesis H_2 , and k is a constant determined so that the critical region should have the required size.

Denote the determinant value $|\sigma_{ij}|$ of the matrix $\|\sigma_{ij}\|$ by σ^2 . Then

$$(2) \quad p_1(z_1, \dots, z_p) = \frac{1}{(2\pi)^{p/2} \sigma} e^{-\frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \sigma^{ij} (z_i - \mu_i)(z_j - \mu_j)},$$

and

$$(3) \quad p_2(z_1, \dots, z_p) = \frac{1}{(2\pi)^{p/2} \sigma} e^{-\frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \sigma^{ij} (z_i - \nu_i)(z_j - \nu_j)},$$

where the matrix $\|\sigma^{ij}\|$ denotes the inverse matrix of the matrix $\|\sigma_{ij}\|$. Taking logarithms of both sides of the inequality (1), we obtain the inequality

$$(4) \quad -\frac{1}{2} \left\{ \sum_i \sum_j \sigma^{ij} [(z_i - \nu_i)(z_j - \nu_j) - (z_i - \mu_i)(z_j - \mu_j)] \right\} \geq \log k.$$

² J. NEYMAN and E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Mem.*, Vol. 1, London, 1936

Multiplying both sides of (4) by 2, we have

$$(5) \quad \sum_j \sum_i \sigma^{ij} [(z_i - \mu_i)(z_j - \mu_j) - (z_i - \nu_i)(z_j - \nu_j)] \geq 2 \log k.$$

The critical region (5) is most powerful with respect to the alternative H_2 , but it cannot be used for our purposes since the parameters $\mu_1, \dots, \mu_p, \nu_1, \dots, \nu_p$ and $\|\sigma_{ij}\|$ are unknown. The optimum estimate of σ_{ij} on the basis of the observations $x_{i\alpha}$ and $y_{i\beta}$ is given by the sample covariance

$$(6) \quad s_{ij} = \frac{\sum_{\alpha=1}^{N_1} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j) + \sum_{\beta=1}^{N_2} (y_{i\beta} - \bar{y}_i)(y_{j\beta} - \bar{y}_j)}{N_1 + N_2 - 2}$$

where $\bar{x}_i = \frac{\sum_{\alpha} x_{i\alpha}}{N_1}$ and $\bar{y}_i = \frac{\sum_{\beta} y_{i\beta}}{N_2}$. The optimum estimates of μ_i and ν_i are given by \bar{x}_i and \bar{y}_i respectively ($i = 1, \dots, p$). Hence for testing H_1 it seems reasonable to use the statistic R which we obtain from the left hand side of (5) by substituting the optimum estimates for the unknown parameters. Thus R is given by

$$(7) \quad R = \sum_j \sum_i s^{ij} [(z_i - \bar{x}_i)(z_j - \bar{x}_j) - (z_i - \bar{y}_i)(z_j - \bar{y}_j)],$$

where $\|s^{ij}\| = \|s_{ij}\|^{-1}$. The critical region for testing H_1 is given by the inequality

$$(8) \quad R \geq C,$$

where C is a constant determined in such a way that the critical region should have the required size. It is interesting to notice that R is proportional to the difference $T_1^2 - T_2^2$ where T_i ($i = 1, 2$) denotes the generalized Student's ratio³ for testing the hypothesis that the set (z_1, \dots, z_p) is drawn from the population π_i . In our case the statistic T_1 cannot be used for testing H_1 , since T_1 is appropriate for this purpose if the class of alternative hypotheses contains all p -variate normal populations having the same covariance matrix as π_1 . In our case the class of alternatives consists merely of a single alternative, namely, the alternative π_2 .

For the sake of certain simplifications we shall propose the use of a statistic U which differs slightly from the statistic R . In order to obtain U , we consider the inequality (5). Since $\sigma^{ij} = \sigma^{ji}$ this inequality can be reduced to

$$(9) \quad \sum_j \sum_i \sigma^{ij} z_i (z_j - \mu_j) \geq k',$$

where k' denotes a certain constant. The statistic U is obtained from the left hand side of (9) by substituting the optimum estimates for the unknown para-

³ See, in this connection H. HOTELLING, "The generalization of Student's ratio," *Annals of Math. Stat.*, Vol. 2, and R. C. BOSE and S. N. ROY, "The exact distribution of the Studentized D^2 statistic," *Sankhya*, Vol. 3.

meters. Thus

$$(10) \quad U = \sum \sum s^{ij} z_i (\bar{y}_j - \bar{x}_j),$$

and the critical region is given by the inequality

$$(11) \quad U \geq d,$$

where the constant d is chosen so that the critical region should have the required size. The statistic U differs from R merely by a term which does not depend on the quantities z_1, \dots, z_p . If N_1 and N_2 are large the difference $U - R$ is practically constant and therefore the critical regions (8) and (11) are identical. The use of U seems to be as justifiable as that of R and because of certain simplifications we propose the use of the critical region (11).

The statistic U is closely connected with the so called discriminant function⁴ introduced by R. A. Fisher for discriminating between the two populations π_1 and π_2 . The discriminant function D is given by

$$(12) \quad D = b_1 d_1 + b_2 d_2 + \dots + b_p d_p$$

where $d_i = \bar{y}_i - \bar{x}_i$ and the coefficient b_i is proportional to $\sum_{j=1}^p s^{ij} d_j$. The coefficients b_1, \dots, b_p are called the coefficients of the discriminant function. We see that U is proportional to the statistic $\sum_{i=1}^p b_i z_i$ which is obtained from the right hand side of (12) by substituting z_i for d_i .

4. Solution of the problem when N_1 and N_2 are large. Denote by $F(U, N_1, N_2 | \pi_i)$ the cumulative probability distribution of U under the hypothesis that the set (z_1, \dots, z_p) has been drawn from the population π_i ($i = 1, 2$). If N_1 and N_2 approach infinity the distribution $F(U, N_1, N_2 | \pi_i)$ converges to a normal distribution, since the variates s_{ij} , \bar{x}_i and \bar{y}_i converge stochastically to the constants σ_{ij} , μ_i and ν_i respectively ($i, j = 1, \dots, p$). Let us denote $\lim_{N_1 \rightarrow N_2 \rightarrow \infty} F(U, N_1, N_2 | \pi_i)$ by $\Phi(U | \pi_i)$ ($i = 1, 2$). Furthermore denote by α_i the mean value, and by σ_i the standard deviation of the distribution $\Phi(U | \pi_i)$ ($i = 1, 2$). It is obvious that $\sigma_1 = \sigma_2 = \sigma$ (say). It is easy to verify that the variates

$$(13) \quad \bar{\alpha}_1 = \sum \sum s^{ij} \bar{x}_i (\bar{y}_j - \bar{x}_j),$$

$$(14) \quad \bar{\alpha}_2 = \sum \sum s^{ij} \bar{y}_i (\bar{y}_j - \bar{x}_j),$$

$$(15) \quad \begin{aligned} \sigma^2 &= \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p s^{ik} s^{jl} (\bar{y}_k - \bar{x}_k) (\bar{y}_l - \bar{x}_l) s_{ij} \\ &= \sum_{k=1}^p \sum_{l=1}^p s^{kl} (\bar{y}_k - \bar{x}_k) (\bar{y}_l - \bar{x}_l), \end{aligned}$$

converge stochastically to the constants α_1 , α_2 and σ^2 respectively.

⁴ R. A. FISHER, "The statistical utilization of multiple measurements," *Annals of Eugenics*, 1938.

Hence for large values of N_1 and N_2 we can assume that U is normally distributed with mean value $\bar{\alpha}_i$ and standard deviation $\bar{\sigma}$ if the hypothesis H_i ($i = 1, 2$) is true. Thus the critical region for testing H_1 is given by the inequality

$$(16) \quad U \geq \bar{\alpha}_1 + \lambda \bar{\sigma},$$

where the constant λ is chosen in such a way that $\frac{1}{\sqrt{2\pi}} \int_{\lambda}^{\infty} e^{-t^2/2} dt$ is equal to the required size of the critical region.

Finally, some remarks about the proper choice of the size of the critical region may be of interest. Two kinds of error may be committed. H_1 may be rejected when it is true, and H_1 may be accepted when H_2 is true. Suppose that W_1 and W_2 are two positive numbers expressing the importance of an error of the first kind and an error of the second kind respectively. If the purpose of the statistical investigation is given it will usually be possible to determine the values of W_1 and W_2 . We shall deal here with the question of determining the size of the critical region as a function of the weights W_1 and W_2 . Denote by P_i the probability that (16) holds under the assumption that H_i is true ($i = 1, 2$). Then P_1 is the size of the critical region (also the probability of an error of the first kind), and $1 - P_2$ is the probability of an error of the second kind. Both probabilities P_1 and P_2 are functions of λ and are given by the following expressions:

$$(17) \quad P_1 = \frac{1}{\sqrt{2\pi}} \int_{\lambda}^{\infty} e^{-t^2/2} dt,$$

and

$$(18) \quad P_2 = \frac{1}{\sqrt{2\pi}} \int_{((\bar{\alpha}_1 - \bar{\alpha}_2)/\bar{\sigma}) + \lambda}^{\infty} e^{-t^2/2} dt.$$

From (13) and (14) we obtain

$$(19) \quad \bar{\alpha}_2 - \bar{\alpha}_1 = \sum_j \sum_i s^{ij} (\bar{y}_j - \bar{x}_i) (\bar{y}_j - \bar{x}_i).$$

Since the right hand side of (19) is positive definite, we have $\bar{\alpha}_2 > \bar{\alpha}_1$. Hence because of (17) and (18) we also have $P_2 > P_1$. By the risk of committing a certain error we understand the probability of that error multiplied by its weight. Hence the risk of committing an error of the first kind is given by $W_1 P_1$, and the risk of committing an error of the second kind is given by $W_2 (1 - P_2)$. It seems reasonable to choose the value of λ so that the two risks become equal to each other, i.e. such that

$$(20) \quad W_1 P_1 = W_2 (1 - P_2).$$

Hence using (17) and (18) we obtain the following equation in λ

$$(21) \quad W_1 \frac{1}{\sqrt{2\pi}} \int_{\lambda}^{\infty} e^{-t^2/2} dt - W_2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{((\bar{\alpha}_1 - \bar{\alpha}_2)/\bar{\sigma}) + \lambda} e^{-t^2/2} dt = 0.$$

Using a table of the normal distribution, the value of λ which satisfies the equation (21) can easily be found. For $H_1 = H_2$ the solution of (21) is given by

$$\lambda = \frac{\bar{\alpha}_2 - \bar{\alpha}_1}{2\sigma},$$

and the critical region is given by the inequality

$$U > \bar{\alpha}_1 + \lambda\sigma = \bar{\alpha}_1 + \frac{\bar{\alpha}_2 - \bar{\alpha}_1}{2} = \frac{\bar{\alpha}_1 + \bar{\alpha}_2}{2}.$$

5. Some results concerning the exact sampling distribution of the statistic U' . If N_1 and N_2 are not large the solution given in section 4 cannot be used and it is necessary to derive the exact sampling distribution of U' . Let

$$(22) \quad (\bar{y}_i - \mu_i) \sqrt{\frac{N_1 N_2}{N_1 + N_2}} = z'_i \quad (i = 1, \dots, p).$$

Then

$$(23) \quad U = \sqrt{\frac{N_1 + N_2}{N_1 N_2}} \sum_i \sum_j s^{ij} z_i z'_j$$

where the variates z'_1, \dots, z'_p are distributed independently of the set (z_1, \dots, z_p) , the mean value of z'_i is equal to $(\mu_i - \mu_1) \sqrt{\frac{N_1 N_2}{N_1 + N_2}}$ and the covariance between z'_i and z'_j is equal to σ_{ij} . It is known that the set of covariances s_{ij} is distributed independently of the set $(z_1, \dots, z_p, z'_1, \dots, z'_p)$ and therefore the distribution of U remains unchanged if instead of (6) we have

$$(24) \quad s_{ij} = \frac{\sum_{\alpha=1}^n t_{i\alpha}^2}{n} \quad (n = N_1 + N_2 - 2),$$

where the variates $t_{i\alpha}$ are distributed independently of the set $(z_1, \dots, z_p, z'_1, \dots, z'_p)$, have a joint normal distribution with mean values zero, $\sigma_{t_{i\alpha} t_{j\beta}} = \sigma_{ij}$ and $\sigma_{t_{i\alpha} t_{j\beta}} = 0$ if $\alpha \neq \beta$. It is necessary to derive the distribution of U' under both hypotheses H_1 and H_2 . In both cases the mean values of $z_1, \dots, z_p, z'_1, \dots, z'_p$ are not zero. Instead of U we will consider the statistic

$$U' = \sum_{i=1}^p \sum_{j=1}^p s^{ij} z_i z'_j$$

which differs from U only in the proportionality factor $\sqrt{\frac{N_1 + N_2}{N_1 N_2}}$. The distributions of U' under the hypotheses H_1 and H_2 are contained as special cases in the distribution of the statistic

$$(25) \quad V = \sum_i \sum_j s^{ij} t_{i,n+1} t_{j,n+2},$$

where s_{ij} is given by (24) and the joint distribution of the variates $t_{i\beta}$ ($i = 1, \dots, p; \beta = 1, \dots, n+2$) is given by

$$(26) \quad \frac{1}{(2\pi)^{p(n+2)/2} \sigma^{n+2}} e^{-\frac{1}{2} \sum_{i=1}^p \sum_{\beta=1}^{n+2} t_{i\beta} t_{i\beta} + (t_{1,n+1} - \xi_1)(t_{2,n+1} - \xi_2) + \dots + (t_{p,n+2} - \eta_p)(t_{1,n+2} - \eta_1)} \\ \times \prod_{\beta=1}^{n+2} \sum_{i=1}^p dt_{i\beta}.$$

The quantities $\xi_1, \dots, \xi_p, \eta_1, \dots, \eta_p$ are constants and σ^2 denotes the determinant value of the matrix $\|\sigma_{ij}\|$.

We will deal here with the distribution of the statistic V given in (25) under the assumption that the joint distribution of the variates $t_{i\beta}$ ($i = 1, \dots, p, \beta = 1, \dots, n+2$) is given by (26).

In order to derive the distribution of V we shall have to prove several lemmas.

LEMMA 1 Let $\|\lambda_{ij}\|$ ($i, j = 1, \dots, p$) be an arbitrary non-singular matrix, and let

$$t'_{i\beta} = \sum_{j=1}^p \lambda_{ij} t_{j\beta} \quad (i = 1, \dots, p; \beta = 1, \dots, n+2).$$

Let furthermore s'_{ij} be given by

$$s'_{ij} = \frac{\sum_{\alpha=1}^n t'_{i\alpha} t'_{j\alpha}}{n}.$$

Then $\sum_j \sum_i s^{ij} t_{i,n+1} t_{j,n+2} = \sum_j \sum_i s'^{ij} t'_{i,n+1} t'_{j,n+2}$, i.e. the statistic V is invariant under non-singular linear transformations.

PROOF. We obviously have

$$(27) \quad t'_{i,n+1} t'_{j,n+2} = \sum_{k=1}^p \sum_{l=1}^p \lambda_{ik} \lambda_{jl} t_{k,n+1} t_{l,n+2}.$$

Furthermore we have

$$(28) \quad s'_{ij} = \sum_{k=1}^p \sum_{l=1}^p \lambda_{ik} \lambda_{jl} s_{kl}.$$

Hence

$$(29) \quad \|s'_{ij}\| = \|\lambda_{ij}\| \|s_{ij}\| \|\bar{\lambda}_{ij}\|$$

where $\bar{\lambda}_{ij} = \lambda_{ji}$.

From (29) we obtain

$$(30) \quad \|s'^{ij}\| = \|\bar{\lambda}^{ij}\| \|s^{ij}\| \|\lambda^{ij}\|,$$

and therefore

$$(31) \quad s'^{ij} = \sum_{k=1}^p \sum_{l=1}^p \lambda^{ki} \lambda^{lj} s^{kl}.$$

Hence from (27) and (31) we obtain

$$(32) \quad \sum_j \sum_i s^{ij} l'_{i,n+1} l'_{j,n+2} = \sum_j \sum_i \sum_k \sum_l \sum_r \sum_s \lambda^{ki} \lambda^{lj} \lambda^{kl} \lambda_{i,n} \lambda_{j,r} l_{n,n+1} l_{r,n+2}.$$

The coefficient of $l_{n,n+1} l_{r,n+2}$ on the right hand side of (32) is given by

$$(33) \quad \sum_j \sum_i \sum_k \sum_l \lambda^{ki} \lambda^{lj} s^{kl} \lambda_{i,n} \lambda_{j,r} = \sum_k \sum_l \left\{ \left(\sum_i \lambda^{ki} \lambda_{i,n} \right) \left(\sum_j \lambda^{lj} \lambda_{j,r} \right) s^{kl} \right\} = s^{kl}.$$

Lemma 1 follows from (32) and (33).

LEMMA 2. *The distribution of V remains unchanged if we assume that the covariance matrix $\|\sigma_{ij}\|$ is equal to the unit matrix, i.e. the joint distribution of the variates $l_{i\beta}$ ($i = 1, \dots, p$; $\beta = 1, \dots, n+2$) is given by*

$$(34) \quad \frac{1}{(2\pi)^{p(n+2)/2}} e^{-\frac{1}{2} \left[\sum_{i=1}^p \sum_{\alpha=1}^{n+2} l_{i\alpha}^2 - \sum_i (l_{i,n+1} - \rho_i)^2 - \sum_i (l_{i,n+2} - \xi_i)^2 \right]},$$

where the constants ρ_i and ξ_i are functions of the constants $\xi_1, \dots, \xi_p, \eta_1, \dots, \eta_p$ and of the σ_{ij} .

Lemma 2 is an immediate consequence of Lemma 1. Hence we have to derive the distribution of V under the assumption that the variates $l_{i\beta}$ have the joint distribution given in (34).

Let R_i ($i = 1, \dots, p$) be the point of the $n+2$ dimensional Cartesian space with the coordinates $l_{i1}, \dots, l_{i,n+2}$. Let $P = (u_1, \dots, u_{n+2})$ and $Q = (v_1, \dots, v_{n+2})$ be two arbitrary points such that $\sum_{\beta=1}^{n+2} u_\beta v_\beta = 0$ and $\sum_{\beta=1}^{n+2} u_\beta^2 = \sum_{\beta=1}^{n+2} v_\beta^2 = 1$.

Denote by O the origin of the coordinate system and let $\bar{l}_{i,n+1}$ be the projection of the vector OR_i on the vector OP . We have

$$(35) \quad \bar{l}_{i,n+1} = \sum_{\beta=1}^{n+2} l_{i\beta} u_\beta \quad (i = 1, \dots, p).$$

Similarly, the projection $\bar{l}_{i,n+2}$ of the vector OR_i on OQ is given by

$$(36) \quad \bar{l}_{i,n+2} = \sum_{\beta=1}^{n+2} l_{i\beta} v_\beta.$$

Let \bar{R}_i ($i = 1, \dots, p$) be the projection of the point R_i on the n -dimensional hyperplane through O and perpendicular to the vectors OP and OQ . Denote the coordinates of \bar{R}_i by $r_{i1}, \dots, r_{i,n+2}$ respectively and let \bar{s}_{ij} be defined by

$$(37) \quad \bar{s}_{ij} = \frac{\sum_{\beta=1}^{n+2} r_{i\beta} r_{j\beta}}{n}.$$

If we rotate the coordinate system so that the $(n+1)$ -axis coincides with OP and the $(n+2)$ -axis coincides with OQ , and if $\bar{l}_{i1}, \dots, \bar{l}_{i,n+2}$ denote the coordinates of R_i ($i = 1, \dots, p$) referred to the new system, then we have

$$(38) \quad \bar{s}_{ij} = \frac{1}{n} \sum_{\beta=1}^{n+2} r_{i\beta} r_{j\beta} = \frac{1}{n} \sum_{\alpha=1}^n \bar{l}_{i\alpha} \bar{l}_{j\alpha}, \text{ and}$$

$$(39) \quad \sum_{\beta=1}^{n+2} l_{i\beta} l_{j\beta} = \sum_{\beta=1}^{n+2} \bar{l}_{i\beta} \bar{l}_{j\beta}.$$

From (38) and (39) we obtain

$$(40) \quad \bar{s}_{ij} = \frac{\sum_{\beta=1}^{n+2} t_{i\beta} t_{j\beta} - \bar{l}_{i,n+1} \bar{l}_{j,n+1} - \bar{l}_{i,n+2} \bar{l}_{j,n+2}}{n}.$$

We will now prove

LEMMA 3. Let \bar{V} be defined by

$$(41) \quad \bar{V} = \sum_i \sum_j \bar{s}_{ij} \bar{l}_{i,n+1} \bar{l}_{j,n+2},$$

where $\bar{l}_{i,n+1}$, $\bar{l}_{i,n+2}$ and \bar{s}_{ij} are given by the formulas (35), (36) and (40) respectively. Let furthermore the joint probability distribution of the variates $t_{i\beta}$ ($i = 1, \dots, p$; $\beta = 1, \dots, n+2$) be given by

$$(42) \quad \frac{1}{(2\pi)^{p(n+2)/2}} e^{-\frac{1}{2} \left[\sum_{i=1}^p \sum_{\beta=1}^{n+2} (t_{i\beta} - u_{i\beta} - t_{i,n+1} v_{i\beta})^2 \right]} \prod_i \prod_{\beta} dt_{i\beta}.$$

Then the distribution of \bar{V} calculated under the assumption that the quantities u_1, \dots, u_{n+2} , v_1, \dots, v_{n+2} are constants and the joint probability distribution of the variates $t_{i\beta}$ is given by (42), is the same as the distribution of V calculated under the assumption that the joint probability distribution of the variates $t_{i\beta}$ is given by (34).

PROOF. If we rotate the coordinate system so that the $(n+1)$ -axis coincides with OP and the $(n+2)$ -axis coincides with OQ , and if $\bar{l}_{i1}, \dots, \bar{l}_{i,n+2}$ denote the coordinates of R_i ($i = 1, \dots, p$) in the new system, then $\bar{l}_{i,n+1}$ and $\bar{l}_{i,n+2}$ are given by the right hand sides of (35) and (36) respectively. Furthermore

$$\bar{s}_{ij} = \frac{\sum_{\alpha=1}^n \bar{l}_{i\alpha} \bar{l}_{j\alpha}}{n}.$$

Hence the distribution of \bar{V} is certainly the same as that of V if the joint probability distribution of the variates $\bar{l}_{i\beta}$ ($i = 1, \dots, p$; $\beta = 1, \dots, n+2$) is given by the expression which we obtain from (34) by substituting $\bar{l}_{i\beta}$ for $t_{i\beta}$. Thus, in order to prove Lemma 3 we have merely to show that if the variates $\bar{l}_{i\beta}$ have the joint probability distribution (34), the variates $t_{i\beta}$ have the joint probability distribution (42). Since the variates $t_{i1}, \dots, t_{i,n+2}$ are obtained by an orthogonal transformation of the variates $\bar{l}_{i1}, \dots, \bar{l}_{i,n+2}$, it follows that the variates $t_{i\beta}$ ($i = 1, \dots, p$; $\beta = 1, \dots, n+2$) are independently and normally distributed with unit variances. We have

$$(43) \quad t_{i\beta} = \sum_{\gamma=1}^{n+2} \lambda_{\beta\gamma} \bar{l}_{i\gamma}$$

where $\lambda_{\beta\gamma}$ is equal to the cosine of the angle between the β -th axis of the original system and γ -th axis of the new system. Since

$$\lambda_{\beta,n+1} = u_{\beta} \quad \text{and} \quad \lambda_{\beta,n+2} = v_{\beta},$$

and since $E(\bar{l}_{i\gamma}) = 0$ for $\gamma = 1, \dots, n$, $E(\bar{l}_{i, n+1}) = \rho$, and $E(\bar{l}_{i, n+2}) = \bar{\epsilon}_i$, it follows from (43) that

$$(44) \quad E(l_{i\beta}) = \rho u_{i1} + \bar{\epsilon}_i x_{i\beta}.$$

Hence Lemma 3 is proved.

We will now prove

LEMMA 4. Let P be a point with the coordinates u_1, \dots, u_{n+2} and Q a point with the coordinates v_1, \dots, v_{n+2} such that $\sum u_{i\beta}^2 = 0$ and $\sum v_{i\beta}^2 = \sum u_{i\beta}^2 = 1$. Denote by L_p the flat space determined by the vectors OR_1, \dots, OR_p ($R_1 = (l_{11}, \dots, l_{1, n+2})$) and let \bar{P} be the projection of P on L_p and \bar{Q} the projection of Q on L_p . Denote furthermore by θ_1 the angle between the vectors OP and $O\bar{P}$, by θ'_1 the angle between OP and $O\bar{Q}$, by θ_2 the angle between OQ and $O\bar{Q}$, by θ'_2 the angle between OQ and $O\bar{P}$, and finally by θ_3 the angle between $O\bar{P}$ and $O\bar{Q}$. Then the statistic \bar{V} defined in (11) is equal to

$$(45) \quad \bar{V} = - \frac{\begin{vmatrix} 0 & a_1 & a_2 \\ b_1 & a_{11} & a_{12} \\ b_2 & a_{12} & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}},$$

where

$$(46) \quad a_1 = \cos^2 \theta_1; \quad a_2 = \cos \theta'_1 \cos \theta_2; \quad b_1 = \cos \theta_1 \cos \theta'_2; \quad b_2 = \cos^2 \theta_2;$$

$$(47) \quad a_{11} = \frac{\cos^2 \theta_1 - a_1^2 - b_1^2}{n}, \quad a_{22} = \frac{\cos^2 \theta_2 - a_2^2 - b_2^2}{n}$$

$$\text{and } a_{12} = \frac{\cos \theta_1 \cos \theta_2 \cos \theta_3 - a_1 a_2 - b_1 b_2}{n}.$$

PROOF. If we rotate the coordinate system in such a way that the $(n+1)$ -axis coincides with OP and the $(n+2)$ -axis coincides with OQ , and if $\bar{l}_{11}, \dots, \bar{l}_{i, n+2}$ are the coordinates of R_i in the new system, then

$$\bar{s}_{ij} = \frac{\sum_{\alpha=1}^n \bar{l}_{i\alpha} \bar{l}_{j\alpha}}{n}.$$

According to Lemma 1 the statistic \bar{V} is invariant under linear transformations of the variables $l_{i\beta}$. Hence \bar{V} is also invariant under linear transformations of the variables $\bar{l}_{i\beta}$. Thus the value of \bar{V} remains unchanged if the points R_1, \dots, R_p are replaced by arbitrary points R'_1, \dots, R'_p of L_p subject to the condition that the vectors OR'_1, \dots, OR'_p be linearly independent. Hence we may assume that the vectors OR_1, \dots, OR_p are perpendicular to each other and lie in the intersection of L_p with the n -dimensional flat space which goes through O and is perpendicular to OP and OQ . Furthermore we may assume that $R_1 = \bar{P}$ and $R_2 = \bar{Q}$. Then OR_i is perpendicular to OP , OQ , OR_1 and OR_2 ($i = 3, \dots, p$).

The statistic \bar{V} can obviously be written in the form:

$$(48) \quad \bar{V} = - \frac{\begin{vmatrix} 0 & \bar{l}_{1,n+1} & \cdots & \bar{l}_{p,n+1} \\ \bar{l}_{1,n+2} & \bar{s}_{11} & \cdots & \bar{s}_{1p} \\ \vdots & \vdots & & \vdots \\ \bar{l}_{p,n+2} & \bar{s}_{p1} & \cdots & \bar{s}_{pp} \end{vmatrix}}{\begin{vmatrix} \bar{s}_{11} & \cdots & \bar{s}_{1p} \\ \vdots & & \vdots \\ \bar{s}_{p1} & \cdots & \bar{s}_{pp} \end{vmatrix}}.$$

Because of our choice of the points R_1, \dots, R_p , we have

$$(49) \quad \bar{l}_{i,n+1} = \bar{l}_{i,n+2} = 0 \quad (i = 3, \dots, p)$$

and

$$(50) \quad \sum_{\beta=1}^{n+2} \bar{l}_{i\beta} \bar{l}_{j\beta} = 0 \quad \text{if } i \neq j \quad (i = 3, \dots, p, j = 1, \dots, p).$$

From (49) and (50) it follows that $\bar{s}_{i,j} = 0$ for $i \neq j$ except \bar{s}_{12} which is not necessarily zero. Hence \bar{V} reduces to the expression

$$(51) \quad \bar{V} = - \frac{\begin{vmatrix} 0 & \bar{l}_{1,n+1} & \bar{l}_{2,n+1} \\ \bar{l}_{1,n+2} & \bar{s}_{11} & \bar{s}_{12} \\ \bar{l}_{2,n+2} & \bar{s}_{12} & \bar{s}_{22} \end{vmatrix}}{\begin{vmatrix} \bar{s}_{11} & \bar{s}_{12} \\ \bar{s}_{12} & \bar{s}_{22} \end{vmatrix}}.$$

We obviously have $\bar{l}_{1,n+1} = a_1$, $\bar{l}_{2,n+1} = a_2$, $\bar{l}_{1,n+2} = b_1$ and $\bar{l}_{2,n+2} = b_2$.

For any two points A and B denote the length of the vector AB by \overline{AB} . Since $n\bar{s}_{11} + (\bar{l}_{1,n+1})^2 + (\bar{l}_{1,n+2})^2 = \overline{OP}^2$, $n\bar{s}_{22} + (\bar{l}_{2,n+1})^2 + (\bar{l}_{2,n+2})^2 = \overline{OQ}^2$ and $n\bar{s}_{12} + \bar{l}_{1,n+1}\bar{l}_{2,n+1} + \bar{l}_{1,n+2}\bar{l}_{2,n+2} = \overline{OP} \cdot \overline{OQ} \cdot \cos \theta_3$, we can easily verify that $\bar{s}_{11} = a_{11}$, $\bar{s}_{12} = a_{12}$ and $\bar{s}_{22} = a_{22}$. Hence Lemma 4 is proved.

The angles θ'_1 and θ'_2 can be expressed in terms of the angles θ_1 , θ_2 and θ_3 . In order to show this, let us rotate the coordinate system so that the first p coordinates lie in the flat space L_p defined in Lemma 4. Let u'_1, \dots, u'_{n+2} be the coordinates of P and v'_1, \dots, v'_{n+2} the coordinates of Q referred to the new axes. Then, since $\overline{OP} = \overline{OQ} = 1$, we have

$$\begin{aligned} \cos \theta_1 &= \sqrt{u_1'^2 + \cdots + u_p'^2}; & \cos \theta'_1 &= \frac{u'_1 v'_1 + \cdots + u'_p v'_p}{\sqrt{v_1'^2 + \cdots + v_p'^2}}; \\ \cos \theta_2 &= \sqrt{v_1'^2 + \cdots + v_p'^2}; & \cos \theta'_2 &= \frac{u'_1 v'_1 + \cdots + u'_p v'_p}{\sqrt{u_1'^2 + \cdots + u_p'^2}}; \end{aligned}$$

$$\text{and} \quad \cos \theta_3 = \frac{u'_1 v'_1 + \cdots + u'_p v'_p}{\sqrt{u_1'^2 + \cdots + u_p'^2} \sqrt{v_1'^2 + \cdots + v_p'^2}}.$$

Hence

$$\cos \theta'_1 = \cos \theta_1 \cos \theta_3 \quad \text{and} \quad \cos \theta'_2 = \cos \theta_2 \cos \theta_3.$$

Introducing the notations

$$m_1 = \cos^2 \theta_1, \quad m_2 = \cos^2 \theta_2 \quad \text{and} \quad m_3 = \cos \theta_1 \cos \theta_2 \cos \theta_3,$$

we have

$$\left\{ \begin{array}{l} a_1 = m_1, \quad a_2 = m_2, \quad b_1 = m_3, \quad b_2 = m_2; \\ a_{11} = \frac{m_1 - m_1^2 - m_2^2}{n}, \quad a_{12} = \frac{m_3(1 - m_1 - m_2)}{n} \\ \text{and} \quad a_{22} = \frac{m_2 - m_2^2 - m_3^2}{n} \end{array} \right.$$

Substituting the above values in (45) we obtain

$$\begin{aligned} \bar{V} &= -n \frac{m_3}{m_1^2 - 1 + m_1 + m_2 - m_1 m_2} \\ &= -n \frac{\cos \theta_1 \cos \theta_2 \cos \theta_3}{\cos^2 \theta_1 \cos^2 \theta_2 \cos^2 \theta_3 - \sin^2 \theta_1 \sin^2 \theta_2}. \end{aligned}$$

Hence, Lemma 4 can be written as

LEMMA 4'. Let P be a point with the coordinates u_1, \dots, u_{n+2} and Q a point with the coordinates v_1, \dots, v_{n+2} . Denote by L_p the flat space determined by the vectors OR_1, \dots, OR_p and let \bar{P} be the projection of P on L_p and \bar{Q} the projection of Q on L_p . Denote furthermore by θ_1 the angle between OP and $O\bar{P}$, by θ_2 the angle between OQ and $O\bar{Q}$ and by θ_3 the angle between $O\bar{P}$ and $O\bar{Q}$. Then the statistic \bar{V} defined in (41) is equal to

$$(45') \quad \bar{V} = -n \frac{\cos \theta_1 \cos \theta_2 \cos \theta_3}{\cos^2 \theta_1 \cos^2 \theta_2 \cos^2 \theta_3 - \sin^2 \theta_1 \sin^2 \theta_2}.$$

If P is a point of the $(n+1)$ -axis and Q a point of the $(n+2)$ -axis, then \bar{V} is identical with the statistic V given in (25). Hence we obtain the following

Geometric interpretation of the statistic V defined in (25). If θ_1 denotes the angle between the $(n+1)$ -axis and the flat space L_p determined by the vectors OR_1, \dots, OR_p , θ_2 the angle between the $(n+2)$ -axis and the flat space L_p , and if θ_3 denotes the angle between the projections of the last two coordinate axes on L_p , then the statistic V is equal to the right hand side of (45').

Denote by S the $2n+1$ -dimensional surface in the $2n+4$ -dimensional space of the variables $u_1, \dots, u_{n+2}, v_1, \dots, v_{n+2}$ defined by the following equations

$$(52) \quad \sum_{\beta=1}^{n+2} u_{\beta}^2 = \sum_{\beta=1}^{n+2} v_{\beta}^2 = 1; \quad \sum_{\beta=1}^{n+2} u_{\beta} v_{\beta} = 0.$$

denote by C the $2n+1$ -dimensional volume of the surface S , i.e.

$$(53) \quad C = \int_S dS.$$

Now we will assume that $u_1, \dots, u_{n+2}, v_1, \dots, v_{n+2}$ are random variables and the joint probability distribution function is defined as follows: the point $(u_1, \dots, u_{n+2}, v_1, \dots, v_{n+2})$ is restricted to points of S and the probability density function of S is defined by

$$(54) \quad \frac{dS}{C}.$$

Hence for any subset A of S the probability of A is equal to the $2n + 1$ -dimensional volume of A divided by the $2n + 1$ -dimensional volume of S . It should be remarked that the probability density function (54) is identical with the probability density function we would obtain if we were to assume that $u_1, \dots, u_{n+2}, v_1, \dots, v_{n+2}$ are independently, normally distributed with zero means and unit variances and calculate the conditional density function under the restriction that $(u_1, \dots, u_{n+2}, v_1, \dots, v_{n+2})$ is a point of S .

LEMMA 5. *The probability distribution of \bar{V} defined in (41), calculated under the assumption that the joint probability density of the variables $u_1, \dots, u_{n+2}, v_1, \dots, v_{n+2}, t_{i\beta}$ ($i = 1, \dots, p; \beta = 1, \dots, n + 2$) is given by the product of (54) and (42), is the same as the distribution of the statistic V calculated under the assumption that the variables $t_{i\beta}$ have the joint probability density function given in (34).*

Lemma 5 is an immediate consequence of lemma 3.

LEMMA 6. *Let L_p be an arbitrary p -dimensional flat space in the $n + 2$ dimensional Cartesian space, and let M_p be the flat space determined by the first p coordinate axes. Assuming that the joint probability density function of $u_\beta, v_\beta, t_{i\beta}$ ($i = 1, \dots, p; \beta = 1, \dots, n + 2$) is given by the product of (54) and (42), the conditional distribution of \bar{V} calculated under the restriction that the points R_1, \dots, R_p lie in L_p , is the same as the conditional distribution of \bar{V} calculated under the restriction that the points R_1, \dots, R_p lie in M_p . The point R_i denotes the point with the coordinates $t_{i1}, \dots, t_{i,n+2}$.*

PROOF. Let P be the point with the coordinates u_1, \dots, u_{n+2} and let Q be the point with the coordinates v_1, \dots, v_{n+2} . Let us rotate the coordinate system so that the first p axes lie in the flat space L_p . Denote the coordinates of P in the new system by u'_1, \dots, u'_{n+2} , those of Q by v'_1, \dots, v'_{n+2} , and those of R_i by $t'_{i1}, \dots, t'_{i,n+2}$ ($i = 1, \dots, p$). Let S' be the surface defined by

$$(55) \quad \sum u'^2_\beta = \sum v'^2_\beta = 1 \quad \text{and} \quad \sum u'_\beta v'_\beta = 0.$$

It is clear that the surface S' is identical with the surface S defined in (52). It is furthermore clear that if the joint density function of $u_1, \dots, u_{n+2}, v_1, \dots, v_{n+2}$ is given by $\frac{dS}{C}$, the joint density function of $u'_1, \dots, u'_{n+2}, v'_1, \dots, v'_{n+2}$

is the same, i.e. it is given by $\frac{dS'}{C}$. It can readily be seen that for any given set of values $u'_1, \dots, u'_{n+2}, v'_1, \dots, v'_{n+2}$ the conditional joint probability density of the variates $t'_{i\beta}$ is given by the function obtained from (42) by substituting

$t'_{i\beta}$ for $t_{i\beta}$, u'_β for u_β and v'_β for v_β , provided that for any given set of values u_1, \dots, u_{n+2} , v_1, \dots, v_{n+2} the joint conditional distribution of the variates $t_{i\beta}$ is given by (42). Hence, if the joint distribution of u_1, \dots, u_{n+2} , v_1, \dots, v_{n+2} and $t_{i\beta}$ ($i = 1, \dots, p$; $\beta = 1, \dots, n+2$) is given by the product of (54) and (42), the joint probability density function of the variates u'_β , v'_β , $t'_{i\beta}$ ($i = 1, \dots, p$; $\beta = 1, \dots, n+2$) is obtained from that of u_β , v_β , $t_{i\beta}$ by substituting S' for S and $t'_{i\beta}$ for $t_{i\beta}$.

According to Lemma 4', \bar{V} can be expressed as a function of the angles θ_1 , θ_2 and θ_3 defined in Lemma 4'. Each angle θ_k ($k = 1, 2, 3$) can be expressed as a function of the variables $t_{i\beta}$, u_β , v_β . It is obvious that the value of θ_k remains unchanged if we substitute $t'_{i\beta}$ for $t_{i\beta}$, u'_β for u_β and v'_β for v_β . Hence also the value of \bar{V} remains unchanged if we substitute $t'_{i\beta}$ for $t_{i\beta}$, u'_β for u_β and v'_β for v_β . Lemma 6 is a consequence of this fact and of the fact that the joint probability density of the variates $t'_{i\beta}$, u'_β and v'_β is identical with that of the variates $t_{i\beta}$, u_β and v_β .

LEMMA 7. Assuming that the joint probability distribution of the variates u_β , v_β , $t_{i\beta}$ ($i = 1, \dots, p$; $\beta = 1, \dots, n+2$) is given by the product of (54) and (42), the conditional joint probability distribution of u_1, \dots, u_{n+2} , v_1, \dots, v_{n+2} , calculated under the restriction that the points $R_i = (t_{i1}, \dots, t_{i,n+2})$ ($i = 1, \dots, p$) lie in the flat space determined by the first p coordinate axes, is given by

$$(56) \quad \frac{e^{-\frac{1}{2} \sum_{\beta=1}^{n+2} \sum_{i=1}^p (\rho_i u_\beta + t_{i\beta} v_\beta)^2}}{\int_S e^{-\frac{1}{2} \sum_{\beta=1}^{n+2} \sum_{i=1}^p (\rho_i u_\beta + t_{i\beta} v_\beta)^2}} f(u_1, \dots, u_{n+2}, v_1, \dots, v_{n+2}) dS,$$

where S denotes the surface defined in (52), and $f(u_1, \dots, u_{n+2}, v_1, \dots, v_{n+2})$ denotes the expected value of

$$(57) \quad \left[\begin{array}{ccc} r_{11} & \dots & r_{1p} \\ r_{21} & \dots & r_{2p} \\ \vdots & & \vdots \\ r_{p1} & \dots & r_{pn} \end{array} \right]^{\frac{n+2-p}{2}} \quad \left(r_{ij} = \sum_{\alpha=1}^n t_{i\alpha} t_{j\alpha} \right)$$

calculated under the assumption that the joint distribution of the variates $t_{i\beta}$ is given by (42).

PROOF. Denote by \bar{R}_i the projection of R_i on the flat space determined by the first p coordinate axes, i.e. $\bar{R}_i = (t_{i1}, \dots, t_{ip}, 0, \dots, 0)$. Let l_i be the length of \bar{R}_i , and let l_i be the distance of \bar{R}_i from the flat space determined by the vectors $0\bar{R}_1, \dots, 0\bar{R}_{i-1}$ ($i = 2, \dots, p$). Then, as is known,

$$(58) \quad l_1 l_2 \dots l_i = \sqrt{\begin{vmatrix} r_{11} & \dots & r_{1i} \\ r_{21} & \dots & r_{2i} \\ \vdots & & \vdots \\ r_{i1} & \dots & r_{ii} \end{vmatrix}} \quad (i = 1, \dots, p),$$

where $r_{kl} = \sum_{\alpha=1}^n t_{k\alpha} t_{l\alpha}$.

We introduce the new variables

$$(59) \quad t_{i\gamma}^* = \frac{l_{i\gamma}}{l_i} \quad (i = 1, \dots, p, \gamma = p+1, \dots, n+2).$$

Then the joint probability density function of the variates u_β , v_β , $t_{i\alpha}$, $t_{i\gamma}^*$ ($i = 1, \dots, p$; $\beta = 1, \dots, n+2$, $\alpha = 1, \dots, p$, $\gamma = p+1, \dots, n+2$) is given by

$$(60) \quad \frac{(l_1 \dots l_p)^{n+2-p}}{C(2\pi)^{p(n+2)/2}} e^{-\frac{1}{2} \left[\sum_{\alpha=1}^p \sum_{\beta=1}^{n+2} (t_{i\alpha} - \rho_{i\alpha} - t_{i\gamma} v_\alpha)^2 + \sum_{\gamma=p+1}^{n+2} \sum_{i=1}^p (t_{i\gamma} - \rho_{i\gamma} - t_{i\alpha} v_\gamma)^2 \right]} \\ \times \left(\prod_i \prod_\alpha dt_{i\alpha} \right) \left(\prod_i \prod_\gamma dt_{i\gamma}^* \right) dS.$$

Substituting zero for $t_{i\gamma}^*$ ($i = 1, \dots, p$, $\gamma = p+1, \dots, n+2$) in (60), we obtain an expression which is proportional to the conditional joint probability density of the variates u_β , v_β , $t_{i\alpha}$ ($\beta = 1, \dots, n+2$, $i = 1, \dots, p$, $\alpha = 1, \dots, p$), calculated under the restriction that the points R_i ($i = 1, \dots, p$) fall in the flat space determined by the first p coordinate axes. Hence this conditional density function is given by

$$(61) \quad A e^{-\frac{1}{2} \sum_{\gamma=p+1}^{n+2} \sum_{\alpha=1}^p (\rho_{i\gamma} + t_{i\alpha} v_\gamma)^2} (l_1 l_2 \dots l_p)^{n+2-p} \\ \times e^{-\frac{1}{2} \left[\sum_{\alpha=1}^p \sum_{\beta=1}^{n+2} (t_{i\alpha} - \rho_{i\alpha} - t_{i\gamma} v_\alpha)^2 \right]} dS \prod_i \prod_\alpha dt_{i\alpha}$$

where A denotes a constant. The conditional distribution of the variates u_β , v_β ($\beta = 1, \dots, n+2$) is obtained from (61) by integrating it with respect to the variables $t_{i\alpha}$ ($i = 1, \dots, p$; $\alpha = 1, \dots, p$). Because of (58), we see that the resulting formula is identical with (56). Hence Lemma 7 is proved.

LEMMA 8. Let $m_1 = u_1^2 + \dots + u_p^2$; $m_2 = v_1^2 + \dots + v_p^2$, and $m_3 = u_1 v_1 + \dots + u_p v_p$. If the joint distribution of the variates u_1, \dots, u_{n+2} , v_1, \dots, v_{n+2} is given by (54), then the joint distribution of m_1, m_2, m_3 is given by

$$(62) \quad \frac{B}{\sqrt{m_1 m_2 (1-m_1)(1-m_2)}} F_p(m_1) F_p(m_2) \Phi_p \left(\frac{m_3}{\sqrt{m_1 m_2}} \right) F_{n+2+p}(1-m_1) \\ \times F_{n+2-p}(1-m_2) \Phi_{n+2-p} \left(\frac{m_3}{\sqrt{(1-m_1)(1-m_2)}} \right) dm_1 dm_2 dm_3$$

where B denotes a constant,

$$(63) \quad F_k(t) = \frac{1}{2^{k/2} \Gamma\left(\frac{k}{2}\right)} t^{(k-2)/2} e^{-t/2} \quad \text{and} \quad \Phi_k(t) = \frac{\Gamma\left(\frac{k}{2}\right)}{\sqrt{\pi} \Gamma\left(\frac{k-1}{2}\right)} (1-t^2)^{(k-3)/2}.$$

PROOF. Let $m'_1 = u_{p+1}^2 + \dots + u_{n+2}^2$, $m'_2 = v_{p+1}^2 + \dots + v_{n+2}^2$,

$m'_3 = u_{p+1}v_{p+1} + \cdots + u_{n+2}v_{n+2}$, $\bar{m}_3 = \frac{m_1}{\sqrt{m_1 m_2}}$ and $\bar{m}'_3 = \frac{m'_1}{\sqrt{m'_1 m'_2}}$. First we calculate the joint distribution of $m_1, m_2, \bar{m}_3, m'_1, m'_2, \bar{m}'_3$ under the assumption that $u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2}$ are normally independently distributed with zero means and unit variances. This joint distribution is given by

$$(64) \quad F_p(m_1)F_p(m_2)\Phi_p(\bar{m}_3)F_{n+2-p}(m'_1)F_{n+2-p}(m'_2) \\ \times \Phi_{n+2-p}(\bar{m}'_3) dm_1 dm_2 d\bar{m}_3 dm'_1 dm'_2 d\bar{m}'_3.$$

Hence the joint distribution of $m_1, m_2, m_3, m'_1, m'_2, m'_3$ is given by

$$(65) \quad \frac{1}{\sqrt{m_1 m_2 m'_1 m'_2}} F_p(m_1)F_p(m_2)\Phi_p\left(\frac{m_3}{\sqrt{m_1 m_2}}\right) F_{n+2-p}(m'_1)F_{n+2-p}(m'_2) \\ \times \Phi_{n+2-p}\left(\frac{m'_3}{\sqrt{m'_1 m'_2}}\right) dm_1 dm_2 dm_3 dm'_1 dm'_2 dm'_3.$$

The required conditional distribution of m_1, m_2, m_3 is equal to the conditional distribution of m_1, m_2, m_3 obtained from the joint distribution (65) under the restrictions $m_1 + m'_1 = 1$, $m_2 + m'_2 = 1$ and $m_3 + m'_3 = 0$. Hence if in (65) we substitute $1 - m_1$ for m'_1 , $1 - m_2$ for m'_2 and $-m_3$ for m'_3 we obtain an expression proportional to the conditional distribution of m_1, m_2, m_3 . This proves Lemma 8.

LEMMA 9. For any point $(u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2})$ of the surface S defined in (52) the expected value of (57) (calculated under the assumption that (42) is the joint distribution of $t_{i\beta}$) is a function of m_1, m_2 , and m_3 only, where m_1, m_2 and m_3 are defined in Lemma 8.

PROOF. Let $\|\lambda_{\alpha\beta}\|$ ($\alpha, \beta = 1, \cdots, p$) be an orthogonal matrix such that

$$(66) \quad \lambda_{1\beta} = \frac{u_\beta}{\sqrt{u_1^2 + \cdots + u_p^2}} \quad (\beta = 1, \cdots, p)$$

and

$$(67) \quad \lambda_{2\beta} = \frac{u_\beta + \lambda v_\beta}{\sqrt{\sum_{\beta=1}^p (u_\beta + \lambda v_\beta)^2}} \quad (\beta = 1, \cdots, p)$$

where

$$\lambda = \frac{-\sum_{\beta=1}^p u_\beta^2}{\sum_{\beta=1}^p u_\beta v_\beta}.$$

Let

$$(68) \quad t'_{i\alpha} = \sum_{\beta=1}^p \lambda_{\alpha\beta} t_{i\beta} \quad (\alpha = 1, \cdots, p).$$

Then the variates $t'_{i\alpha}$ are independently and normally distributed with unit variances. Since for any point of S , $E(t_{i\alpha}) = \rho_i u_\alpha + \xi_i v_\alpha$, we have because of (66), (67) and (68)

$$E(t_{i\gamma}) = 0 \quad (i = 1, \dots, p, \gamma = 3, 4, \dots, p),$$

$$E(t_{i1}) = \varphi_{i1}(m_1, m_2, m_3),$$

$$\text{and} \quad E(t_{i2}) = \varphi_{i2}(m_1, m_2, m_3)$$

Hence the joint distribution of the variates $t'_{i\alpha}$ ($i = 1, \dots, p; \alpha = 1, \dots, p$) depends merely on m_1, m_2 and m_3 . Since $r_{ij} = \sum_{\alpha=1}^p t_{i\alpha} t_{j\alpha} = \sum_{\alpha=1}^p t'_{i\alpha} t'_{j\alpha}$, the expression (57) can be expressed as a function of the variables $t'_{i\alpha}$. Hence the distribution of the expression (57) depends merely on the parameters m_1, m_2 , and m_3 . This proves Lemma 9.

The main result of this section is the following

THEOREM. Let V be the statistic given in (25) and let the joint distribution of the variates $t_{i\beta}$ ($i = 1, \dots, p; \beta = 1, \dots, n+2$) be given by (34). Then the probability distribution of V is the same as the distribution of

$$(69) \quad -n \frac{m_3}{m_3^2 - (1 - m_1)(1 - m_2)}$$

where the joint distribution of m_1, m_2 and m_3 is equal to a constant multiple of the product of the following three factors: the expression (62), the exponential $e^{\lambda(m_1 \sum_{i=1}^p t_{i1}^2 + 2m_3 \sum_{i=1}^p t_{i1} t_{i2} + m_2 \sum_{i=1}^p t_{i2}^2)}$ and the expected value of

$$(70) \quad \left(\begin{matrix} r_{11} & \dots & r_{1p} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ r_{p1} & \dots & r_{pp} \end{matrix} \right)^{(n+2-p)/2} \quad \left(r_{ij} = \sum_{\alpha=1}^p t_{i\alpha} t_{j\alpha} \right).$$

The expected value of (70) is calculated under the assumption that the variates $t_{i\alpha}$ are normally and independently distributed with unit variances and $E(t_{i\alpha}) = \rho_i u_\alpha + \xi_i v_\alpha$ ($i = 1, \dots, p; \alpha = 1, \dots, p$) where $\sum_{\alpha=1}^p u_\alpha^2 = m_1$, $\sum_{\alpha=1}^p v_\alpha^2 = m_2$ and $\sum_{\alpha=1}^p v_\alpha u_\alpha = m_3$. The domain of the variables m_1, m_2 and m_3 is given by the inequalities: $0 \leq m_1 \leq 1; 0 \leq m_2 \leq 1; -\sqrt{m_1 m_2} \leq m_3 \leq \sqrt{m_1 m_2}$.

PROOF. First we note that the expected value of (70) is a function of m_1, m_2 and m_3 only. Let P be the point with the coordinates u_1, \dots, u_{n+2} , and Q the point with the coordinates v_1, \dots, v_{n+2} . Assume that the points $R_i = (t_{i1}, \dots, t_{i,n+2})$ ($i = 1, \dots, p$) lie in the flat space determined by the first p coordinate axes. Assume furthermore that $u_1 v_1 + \dots + u_{n+2} v_{n+2} = 0$ and that the lengths of the vectors OP and OQ are equal to 1. Then

$$\cos \theta_1 = \sqrt{u_1^2 + \dots + u_p^2}; \quad \cos \theta_2 = \sqrt{v_1^2 + \dots + v_p^2}$$

and

$$\cos \theta_3 = \frac{u_1 v_1 + \dots + u_p v_p}{\sqrt{u_1^2 + \dots + u_p^2} \sqrt{v_1^2 + \dots + v_p^2}},$$

where θ_1 denotes the angle between OP and the flat space L_p , determined by the vectors OR_1, \dots, OR_p ; θ_2 denotes the angle between OQ and L_p , and θ_3 denotes the angle between the projections of OP and OQ on L_p . According to Lemma 4' the statistic \bar{V} defined in (41) is equal to

$$\begin{aligned} \bar{V} &= -n \frac{\cos \theta_1 \cos \theta_2 \cos \theta_3}{\cos^2 \theta_1 \cos^2 \theta_2 \cos^2 \theta_3 - \sin^2 \theta_1 \sin^2 \theta_2} \\ (71) \quad &= -n \frac{m_3}{m_3^2 - (1 - m_1)(1 - m_2)} \end{aligned}$$

where

$$\begin{aligned} (72) \quad m_1 &= \cos^2 \theta_1 = u_1^2 + \dots + u_p^2, \quad m_2 = \cos^2 \theta_2 = v_1^2 + \dots + v_p^2 \\ \text{and } m_3 &= \cos \theta_1 \cos \theta_2 \cos \theta_3 = u_1 v_1 + \dots + u_p v_p. \end{aligned}$$

It follows from Lemmas 5 and 6 that the distribution of V is the same as the conditional distribution of \bar{V} calculated under the assumption that the unconditional joint probability density of the variates u_i, v_i and t_{ij} is given by the product of (54) and (42) and under the restriction that the points R_i ($i = 1, \dots, p$) fall in the flat space determined by the first p coordinate axes. Since $e^{-\frac{1}{2} \sum_{i=1}^{n+2} \sum_{j=1}^p (\rho_i u_j + t_{ij} v_j)^2}$ is a constant multiple of

$$(73) \quad e^{\frac{1}{2}(m_1 \sum \rho_i^2 + 2m_3 \sum \rho_i t_{ij} + m_2 \sum t_{ij}^2)}$$

from Lemmas 7, 8 and 9 it follows readily that the joint conditional distribution of $m_1 = u_1^2 + \dots + u_p^2$, $m_2 = v_1^2 + \dots + v_p^2$ and $m_3 = u_1 v_1 + \dots + u_p v_p$ is equal to a constant multiple of the product of (62), (73) and the expected value of 70. This proves our theorem.

It can be shown that the variates m_1, m_2 and m_3 are of the order $\frac{1}{n}$ in the probability sense. Hence

$$(74) \quad -n \frac{m_3}{m_3^2 - (1 - m_1)(1 - m_2)} = nm_3(1 + \epsilon)$$

where ϵ is of the order $\frac{1}{n}$. Hence we can say: *even for moderately large n the distribution of the statistic \bar{V} is well approximated by the distribution of nm_3 , where the joint distribution of m_1, m_2 and m_3 is equal to a constant multiple of the product of (62), (73) and the expected value of (70).*

If $n + 2 - p$ is an even integer, the expected value of (70) is obviously an elementary function of m_1, m_2 and m_3 . Hence, if $n + 2 - p$ is even, the joint distribution of m_1, m_2 and m_3 is also an elementary function of m_1, m_2 and m_3 .

If the constants ρ_i and ζ_i ($i = 1, \dots, p$) in formula (34) are equal to zero, the expected value of (70) is a constant and the joint distribution of m_1, m_2 and m_3 is given by (62).

ASYMPTOTIC DISTRIBUTION OF RUNS UP AND DOWN¹

By J. WOLFOWITZ

Columbia University

1. Introduction. Let a_1, a_2, \dots, a_n be any n unequal numbers and let $S = (h_1, h_2, \dots, h_n)$ be a random permutation of them, with each permutation having the same probability, which is therefore $\frac{1}{n!}$. Let R be the sequence of signs (+ or -) of the differences $h_{i+1} - h_i$ ($i = 1, 2, \dots, n - 1$). Then R is also a chance variable. A sequence of p successive + (-) signs not immediately preceded or followed by a + (-) sign is called a run up (down) of length p . The term "run" applies to both runs up and runs down. As an example, if $S = (4\ 6\ 2\ 3\ 5)$, then in $R = (+\ -\ +\ +)$ there are three runs, one up of length one, one down of length one, and one up of length two.

The purpose of this paper is to establish several theorems about the limiting distributions of a class of functions of runs up and down. These results are applicable to certain techniques which have been employed in quality control and the analysis of economic time series. They are also shown to apply to a large class of "runs."

2. Joint distribution of runs of several lengths. Let r_p be the number of runs of length p in R and r'_p the number of runs of length p or more in R . Then r_p and r'_p are chance variables. The expectations $E(r_p)$ and $E(r'_p)$, the variances $\sigma^2(r_p)$ and $\sigma^2(r'_p)$, and the covariances $\sigma(r_{p_1}r_{p_2})$ are given by Levene and Wolfowitz [1]. They are all of the order n . Let

$$y_p = \frac{r_p - E(r_p)}{\sqrt{n}},$$

$$y'_p = \frac{r'_p - E(r'_p)}{\sqrt{n}}.$$

Our first results are embodied in the following theorem.

THEOREM 1. *Let l be any non-negative integer. The joint distribution of $y_1, \dots, y_l, y'_{l+1}$, approaches the normal distribution as $n \rightarrow \infty$.*

We shall give the proof for the case $l = 1$, but it will easily be seen to be perfectly general.

Let $x_{p,i} = 1$ if the sign (+ or -) of $h_{i+1} - h_i$ is the initial sign of a run of length p , and let $x_{p,i} = 0$ otherwise. Let $w_{p,i} = 1$ if the sign of $h_{i+1} - h_i$ is the

¹ Part of the results of this paper was presented to the Institute of Mathematical Statistics and the American Mathematical Society at their joint meeting in New Brunswick, N. J., on September 13, 1943.

initial sign of a run of length p or more, and let $w_{pn} = 0$ otherwise. Let $x_{pn} = w_{pn} = 0$. Then

$$r_1 = \sum_{i=1}^n x_{1i},$$

$$r'_2 = \sum_{i=1}^n w_{2i}.$$

Now write $\alpha = n^{\frac{1}{2}}$, $\beta = n^{\frac{1}{4}}$, and consider the β sequences

$$h_{(j-1)\alpha+1}, h_{(j-1)\alpha+2}, \dots, h_{j\alpha} \quad (j = 1, 2, \dots, \beta).$$

(Strictly speaking, we should employ the largest integer in α . Since what is meant is clear and since we are dealing with an asymptotic property, we shall omit this useless nicety.) Let x'_{pi} and w'_{pi} have the same definitions relative to each of these sequences that x_{pi} and w_{pi} have relative to the sequence S . The accented and unaccented x 's and w 's are not always the same, because the partitioning of the sequence S sometimes breaks up runs and creates others. Thus we might have $x_{p\alpha} = 1$, but $x'_{p\alpha}$ always $= 0$.

It is easy to see that there exists a positive number d such that

$$\sum_{i=1}^n |x_{1i} - x'_{1i}| < d\beta,$$

$$\sum_{i=1}^n |w_{2i} - w'_{2i}| < d\beta.$$

If, therefore, we define

$$z_1 = \frac{\sum_{i=1}^n [x'_{1i} - E(x'_{1i})]}{\sqrt{n}},$$

$$z'_2 = \frac{\sum_{i=1}^n [w'_{2i} - E(w'_{2i})]}{\sqrt{n}},$$

we have

$$|z_1 - y_1| < \frac{2d\beta}{\sqrt{n}}$$

$$|z'_2 - y'_2| < \frac{2d\beta}{\sqrt{n}}$$

and

$$\frac{d\beta}{\sqrt{n}} \rightarrow 0.$$

Hence, if the joint limiting distribution of z_1 and z'_2 is normal, so is that of y_1 and y'_2 .

The chance variables

$$r_{1j} = \sum_{i=(j-1)\alpha+1}^{j\alpha} x'_{1i},$$

$$r'_{2j} = \sum_{i=(j-1)\alpha+1}^{j\alpha} w'_{2i}, \quad (j = 1, 2, \dots, \beta)$$

have the same joint distribution for all values of j . For x'_{1i} and w'_{2i} , $((j-1)\alpha < i < j\alpha)$, depend only on the relative magnitude of the elements of the sequence

$$h_{(j-1)\alpha+1}, \dots, h_{j\alpha},$$

not upon the particular values which the elements take, and all permutations of the sequence have equal probability. Clearly r_{1j} and r'_{2j} are independent, in the probability sense, of $r_{1j'}$ and $r'_{2j'}$ ($j \neq j'$), because of the definitions of x'_{1i} and w'_{2i} . (However, r_{1j} and r'_{2j} are not independent, because x'_{1i} and w'_{2i} cannot both be 1.) From the results of [1] it follows that for sufficiently large n the absolute value of the correlation coefficient between r_{1j} and r'_{2j} is less than a number smaller than 1. By the methods of [1] it can easily be shown that the ratio of the fourth order moments of r_{1j} and r'_{2j} about their means to the square of the variance of either, is bounded for sufficiently large n . Hence by Liapounoff's theorem (see, for example, Cramer [2], Uspensky [5]), z_1 and z'_2 are jointly normally distributed in the limit. Hence so are y_1 and y'_2 and the theorem is proved.

3. Generalization of Theorem 1. Examination of the proof of Theorem 1 shows that it rests on the following two properties of runs up and down.

a) Partition of the sequence S into subsequences affects at most d runs in each sub-sequence, where d is a fixed positive number independent of n .

b) After partition the totals of runs of each length in any sub-sequence (the definition now relates to the subsequence) are independent in the probability sense of the totals of runs in any other subsequence, and satisfy some condition (such as the Liapounoff) sufficient to make the components of the sum of the vectors jointly normally distributed in the limit.

Hence if we adopt other definitions of runs which meet conditions (a) and (b) above, the total numbers of each of these various kinds of runs will be in general jointly asymptotically normally distributed. For example, if s_p and s'_p be the numbers of runs *up* of length p and of length p or more, respectively, and if t_p and t'_p are the same quantities referring to runs *down*, then, with l and k any positive integers,

$$s_1, s_2, \dots, s_l, \quad s'_{(l+1)}, \quad t_1, t_2, \dots, t_k$$

are jointly asymptotically normally distributed. However, if t'_{k+1} is included in this set, since

$$s'_1 = s_1 + s_2 + \cdots + s_l + s'_{l+1}$$

and

$$t'_1 = t_1 + t_2 + \cdots + t_k + t'_{k+1}$$

differ by at most one, the limiting distribution is degenerate, i.e., its covariance matrix is only semi-definite.

As another example, if we define a bizarre run as, say, the occurrence of a run up of length 5, followed, 17 elements later, by a run down of length 14, then the number of runs of this type is asymptotically normally distributed with expectation and variance of order n .

4. Additive functions of runs of all lengths. Combining the numbers of runs of all lengths greater than a given length generally involves a loss of information. The following theorem on additive functions of runs up and down may be of general interest and of utility in avoiding this undesirable situation.

THEOREM 2. Let $f(i)$ be a function, defined for all positive integral values of i , which fulfills the following conditions:

a) There exists a pair of positive integers, a and b , such that

$$(4.1) \quad \frac{f(a)}{f(b)} \neq \frac{a}{b}$$

b) for any $\epsilon_1 > 0$ there exists a positive integer $N(\epsilon_1)$ such that, for all $n \geq N(\epsilon_1)$,

$$(4.2) \quad \sum_{i=N(\epsilon_1)}^{i=n-1} |f(i)| \sigma(r_i) < \epsilon_1 n$$

where n , of course, has the same meaning as in the preceding sections. Let $F(S)$, a function of the chance sequence S , be defined as follows:

$$(4.3) \quad F(S) = \sum_{i=1}^{(n-1)} f(i)r_i.$$

Then the distribution of $\frac{F(S) - E[F(S)]}{\sigma\{F(S)\}}$ approaches the normal distribution as $n \rightarrow \infty$.

As an example, let $f(i) \equiv 1$. Then $F(S) \equiv r'_1$, whose limiting distribution is normal by Theorem 1.

This theorem is the exact analogue of Theorem 2 of [3] and the proof of the latter carries over without difficult changes except in one important respect. A difficulty in the proof of the theorem in [3] lay in proving Lemma 4, and this lemma has to be proved completely anew. We shall limit ourselves here to doing just that. Lemma 2 of Theorem 2 of [3], whose only role was to help in proving Lemma 4, has no analogue in our present problem, but all the others do. It will therefore be sufficient if we prove the following:

LEMMA. *There exists a constant $c > 0$, such that, for all n sufficiently large,*

$$(4.4) \quad \sigma^2[F(S)] > cn.$$

Condition (a) of the theorem is imposed simply in order that the result be not trivial. For, if (a) does not hold, we have that

$$f(i) \equiv \text{if } (1),$$

and

$$\begin{aligned} F(S) &\equiv f(1) \sum ir_i \\ &\equiv (n-1)f(1) = \text{a constant.} \end{aligned}$$

Suppose that

$$f(i) \equiv ui + v,$$

with u and v constants, and $v \neq 0$. Then by Theorem 1

$$\begin{aligned} F(S) &= u(n-1) + vr'_1 \\ &= vr'_1 + \text{a constant} \end{aligned}$$

is asymptotically normally distributed with variance of order n . Without loss of generality we may therefore assume that

$$(4.5) \quad f(i) \neq ui + v.$$

From this it follows that there exists an integer $A \geq 2$ such that

$$(4.6) \quad f(A-1) + f(A+1) \neq 2f(A).$$

Our object is to prove that $\sigma^2[F(S)]$ is at least of order n . The basic idea of the proof will be to construct two sets, say L_1 and L_2 , of sequences S , such that the (same) probability of each is not less than a positive lower bound independent of n , and such that there exists a one-to-one correspondence between the sequences of L_1 and those of L_2 so that, if S_1 is a member of L_1 and S_2 the corresponding sequence in L_2 ,

$$|F(S_1) - F(S_2)| \geq g\sqrt{n},$$

where g is a positive constant independent of n . It is easy to see that such a construction would prove the lemma.

We shall call the subsequence $(h_i, h_{i+1}, \dots, h_{i+2A})$ of S , a run of type T_1 or simply a run T_1 (the notion will be used only for the proof of this lemma) if the following conditions are fulfilled:

- (4.7) each of the signs of $(h_{i+1} - h_i)$ and $(h_{i+A+1} - h_{i+A})$ is the initial sign of a run of length A .
- (4.8) if $i \neq 1$, the sign of $(h_i - h_{i-1})$ is not the final sign of a run of length A .
- (4.9) if $i + 2A \neq n$, the sign of $(h_{i+2A+1} - h_{i+2A})$ is not the initial sign of a run of length A .

(4.10) after the transformation H , which interchanges $h_{i, i-1}$ and $h_{i, i+1}$, has operated on the run, the sign of $(h_{i, i-1} \dots h_i)$ is the initial sign of a run of length $A - 1$, and the sign of $(h_{i, i+A-1} \dots h_{i+A})$, in the new ordering, is the initial sign of a run of length $A + 1$.

Thus, with $A = 2$ and $n = 7$, if $S = (7145326)$, then $R = (- + - + - + -)$, and $(1\ 4\ 5\ 3\ 2)$ is a run T_1 , for after the transformation H has been applied we have $(1\ 5\ 4\ 3\ 2)$ which gives $(+ - - -)$. The result of the operation H on a run T_1 will be called a run T_2 .

The number r^* of runs T_1 and the number r^{**} of runs T_2 each have expected values and variances of order n , by considerations similar to those of [1]. Hence, for an arbitrarily small positive ϵ there exists a positive constant q such that, for all n sufficiently large, the probability $P\{r^* + r^{**} \geq qn\}$ of the set L^* of sequences S which satisfy the relation in braces, is not less than $1 - \epsilon$.

The set L^* can be divided into disjunct sets (families) as follows: Let $S(0)$ be any sequence S in L^* which has no runs T_2 (any doubt about the existence of such sequences will be soon removed) and let $r^*[S(0)] = m$. Hence $m \geq qn$. Operating with the transformation H on each of the m runs T_1 of $S(0)$ we get a set $S(1)$ of m different sequences for each of which $r^* = m - 1$, $r^{**} = 1$. Operating again with H on each of the pairs of runs T_1 of the sequence $S(0)$ we get a set $S(2)$ of $\binom{m}{2}$ distinct sequences for each of which $r^* = m - 2$, $r^{**} = 2$, etc. The process stops with $S(m)$, which contains a single sequence, for which $r^* = 0$, $r^{**} = m$. The set $S(i)$ contains $\binom{m}{i}$ different sequences for each of which $r^* = m - i$, $r^{**} = i$. The union of the sets $S(i)$ ($i = 1, 2, \dots, m$) will be called the family whose generator is $S(0)$. The sets $S(i)$ are obviously disjunct. Any sequence S in L^* belongs to one and only one family. For if we operate on all of its runs T_2 with H (which is its own inverse), we obtain the generator of the family to which it belongs. This also proves the existence of sequences in L^* for which $r^{**} = 0$.

Consider any family F whose generator is a sequence for which $r^* = m \geq qn$. It is easy to see that, when n is sufficiently large, the ratio of the total number of sequences S in the sets L_1^* and L_2^* , where

$$L_1^* = \sum_{i=0}^{m-\sqrt{m}} S(i),$$

and

$$L_2^* = \sum_{i=\frac{1}{2}(m+\sqrt{m})}^{i=m} S(i),$$

to the total number of sequences in F is greater than a fixed positive constant K' .

We are now ready to construct L_1 and L_2 . The set L_1 is the union of the sets L_1^* of all the families in L^* , and the set L_2 is the union of the sets L_2^* of all the families in L^* . The probability of L_1 and of L_2 is therefore not less than $\frac{1}{2}K'(1 - \epsilon)$. The one-to-one correspondence is effected as follows: The subset $S\left(\frac{m}{2} - \frac{\sqrt{m}}{2} - j\right)$ of the set L_1^* of any family is to correspond to the

subset $S\left(\frac{m}{2} + \frac{\sqrt{m}}{2} + j\right)$ ($j = 0, 1, 2, \dots, \frac{m}{2} - \frac{\sqrt{m}}{2}$) of the set L_2^* of the same family. The individual sequences of either of the two subsets may be made to correspond to those of the other in any manner whatsoever. Any sequence S_1 in L_1 and its corresponding sequence S_2 in L_2 thus differ only in the numbers of runs T_1 and T_2 , but are identical in the numbers of all other runs. They differ in at least \sqrt{m} runs. Hence,

$$\begin{aligned} |F(S_1) - F(S_2)| &\geq \sqrt{m} |2f(A) - f(A-1) - f(A+1)| \\ &\geq \sqrt{qn} |2f(A) - f(A-1) - f(A+1)|. \end{aligned}$$

This is the required result with

$$g = \sqrt{q} |2f(A) - f(A-1) - f(A+1)|.$$

Hence the lemma and the theorem are proved.

The remarks of section 3 also apply to Theorem 2.

5. The distribution of long runs. Certain tests in use in quality control of manufactured products are based on the occurrence of long runs. Since the mean and variance of r_p , for any fixed p , are of order n , it follows that the probability that $r_p \neq 0$ approaches 1 (with increasing n). In order to base a test on the occurrence of a run of length p in long sequences it is therefore necessary to make p a function of n . This function must be a suitable one, because if p is, for example, of the order n , the probability that $r_p = 0$ approaches 1; p should, therefore, be neither too short nor too long.

The following theorem will help give the answer to this problem:

THEOREM 3. *Let p vary with n , so that*

$$\frac{(p+1)!}{n} = \frac{1}{K}$$

with K a fixed positive number. Then

$$\lim_{n \rightarrow \infty} P\{r_p = j\} = e^{-2K} \frac{(2K)^j}{j!} \quad (j = 0, 1, 2, \dots)$$

i.e., r_p has in the limit the Poisson distribution with mean $2K$

The proof will consist in showing that the moments of r_p approach the moments of a Poisson distribution with mean $2K$ as $n \rightarrow \infty$. This is sufficient (v. Mises [4]).

Let $x_i = 1$ if the sign of $h_{i+1} - h_i$ is the initial sign of a run of length p , and $x_i = 0$ otherwise. The probability that $x_i = 1$ is, by [1], Section [4], $\frac{2(p^2 + 3p + 1)}{(p+3)!}$ for all i with a fixed number of exceptions.² Write $B = \frac{2}{(p+1)!}$; then

$$P\{x_i = 1\} = B + o(B),$$

² Since these exceptions (at the ends of the sequence S) have no effect on the asymptotic theory, they will henceforth be ignored

where the symbol $o(B)$ means that $\lim_{B \rightarrow 0} \frac{o(B)}{B} = 0$. Let y_i , $i = 1, 2, \dots, n$ be independent chance variables with the same distribution: $P\{y_i = 1\} = B$, $P\{y_i = 0\} = 1 - B$. Then it is easy to see that $Y = \sum_{i=1}^n y_i$ has in the limit the Poisson distribution with mean $2K$ and that its moments approach the moments of the same Poisson distribution. Hence it will be sufficient to show that in the limit Y and r_p have the same moments.

If $q, \alpha_1, \alpha_2, \dots, \alpha_q$ and $i_1 < i_2 < \dots < i_q$ are positive integers, we have that

$$\begin{aligned} E(y_{i_1}^{\alpha_1} y_{i_2}^{\alpha_2} \cdots y_{i_q}^{\alpha_q}) &= E(y_{i_1} y_{i_2} \cdots y_{i_q}) \\ (5.1) \qquad \qquad \qquad &= \prod_{j=1}^q E(y_{i_j}) = B^q \end{aligned}$$

and

$$(5.2) \qquad 0 \leq E(x_{i_1}^{\alpha_1} x_{i_2}^{\alpha_2} \cdots x_{i_q}^{\alpha_q}) = E(x_{i_1} x_{i_2} \cdots x_{i_q}).$$

Also

$$(5.3) \qquad E(r_p^l) = E\left[\sum_{i=1}^n x_i\right]^l.$$

After expansion of the right member of (5.3), we may replace, in accord with (5.2), each of the non-zero exponents of the x 's by 1. The same operation on the terms of the expansion of the right member of

$$(5.4) \qquad E(Y^l) = E\left[\sum_{i=1}^n y_i\right]^l,$$

is valid in accord with (5.1).

Let $i_1 < i_2 < \dots < i_q$. In the expression

$$(5.5) \qquad E(x_{i_1} x_{i_2} \cdots x_{i_q}),$$

let q be the "weight." A subsequence of consecutive x 's in (5.5) (it may consist of a single x) which is such that the indices of two consecutive x 's differ by less than $(p+3)$, while the subsequence cannot be expanded on either side without violating this requirement, will be called a "cycle." Let c be the number of cycles in (5.5). By [1], Section 4, if x_i and x_j are in different cycles, i.e., $|i - j| \geq (p+3)$, then x_i and x_j are independently distributed. If, therefore, $q = c$, we have that

$$(5.6) \qquad E(x_{i_1} x_{i_2} \cdots x_{i_q}) = \prod_{j=1}^q E(x_{i_j}) = B^q + o(B^q).$$

If $q > c = 1$, we have, also from [1], Section 4, that

$$(5.7) \qquad E(x_{i_1} x_{i_2} \cdots x_{i_q}) \leq E(x_{i_1} x_{i_2}) = o(B).$$

If $q > c$ and if there are two indices in the expression (5.5) which differ by less than p , then

$$(5.8) \quad E(x_{i_1} x_{i_2} \cdots x_{i_q}) = 0.$$

For x_i and x_j cannot both initiate runs of length p if $|i - j| < p$.

Let us now return to the expansions of the right members of (5.3) and (5.4), in which the exponents have been replaced as described before. Let the weight and cycle definitions also apply to terms of the type

$$(5.9) \quad E(y_{i_1} y_{i_2} \cdots y_{i_q}).$$

From (5.1) and (5.6) it follows that, in the limit, the contributions to $E(r_p^l)$ and $E(Y^l)$ of the sums of those terms for which $q = c$, are the same. Let W and W' be the sums of all the remaining terms in $E(r_p^l)$ and $E(Y^l)$, respectively. If we can show that

$$(5.10) \quad \lim W = \lim W' = 0$$

we will have proven that

$$(5.11) \quad \lim E(r_p^l) = \lim E(Y^l)$$

and with it the theorem.

Let $B = O[f(n)]$ mean, as usual, that $|B| \leq Mf(n)$ for all n and a fixed $M > 0$. The number of terms in W' with fixed q and c ($c < q$, by definition of W') is $O(n^c p^{q-c})$. From (5.1) the value of the sum of all such terms is $O(B^q n^c p^{q-c})$. Now

$$nB = O(1)$$

by the hypothesis of the theorem. From the definition of p ,

$$p = o(n)$$

and hence

$$pB = o(1).$$

Therefore

$$\begin{aligned} B^q n^c p^{q-c} &= (nB)^c (pB)^{q-c} \\ &= o(1). \end{aligned}$$

Since $q \leq l$, there are only a fixed number of such sums. Hence $\lim W' = 0$.

The number of terms in W with fixed q and c ($c < q$) is $O(n^c p^{q-c})$. However, most of these are of the type in (5.8) and therefore vanish. Those which do not vanish are $O(n^c)$ in number. Since $q > c$ we have by application of (5.7) that each term is $o(B^c)$. Hence the value of the sum of these terms is $o(n^c B^c) = o(1)$. Since $q \leq l$, there are a fixed number of such sums. Hence $\lim W = 0$.

This proves (5.10) and with it the theorem.

It is possible to generalize this result in a manner similar to that of Section 3.

The author is obliged to W. Allen Wallis who first drew his attention to problems in runs up and down, and to Howard Levene, who read the manuscript of this paper.

REFERENCES

- [1] H. LEVENE and J. WOLFOWITZ, *Annals of Math. Stat.*, Vol. 15 (1944).
- [2] HARALD CRAMÉR, *Random Variables and Probability Distributions*, Cambridge, 1937.
- [3] J. WOLFOWITZ, *Annals of Math. Stat.*, Vol. 13 (1942), p. 247.
- [4] R. v. MISES, *Zeitschrift fuer die angewandte Math. und Mechanik*, Vol. 1 (1921), p. 258.
- [5] J. V. USPENSKY, *Introduction to Mathematical Probability*, New York, 1937.

STATISTICAL ANALYSIS OF CERTAIN TYPES OF RANDOM FUNCTIONS

By H. HURWITZ, JR. AND M. KAC

Cornell University

1. Introduction. In solving certain physical problems (Brownian movements, shot effect) one is often led to the study of superpositions of random pulses. More precisely, one is led to sums of the type

$$(1) \quad F(t) = \sum_{i=1}^N f(t - t_i),$$

where N and the t_i 's are random variables and a function $P(t)$ is given such that $\int_{\Delta} P(t) dt$ represents the average number of pulses occurring during the time interval Δ .

We propose to give a fairly detailed treatment of those statistical properties of $F(t)$ which may be of interest to a physicist and at the same time pay careful attention to the mathematical assumptions which underly the applications. It may also be pointed out that our results could be applied to the theory of time series.

2. Statistical assumptions and the distribution of N . The statistical assumptions can be formulated as follows:

1. The t_i 's form an infinite sequence of independent identically distributed random variables each having $p(t)$ as its probability density.
2. N is capable of assuming the values 0, 1, 2, 3, \dots only, and N is independent of the t_i 's.
3. If $M(\Delta; N)$ denotes the number of those t_i 's among the first N , which fall within the interval Δ , then for non-overlapping intervals Δ_1 and Δ_2 the random variables $M(\Delta_1; N)$ and $M(\Delta_2; N)$ are independent.

We now state our first theorem.¹

THEOREM 1. *Assumptions 1, 2, 3 imply that N is distributed according to Poisson's law, i.e.*

$$\text{Prob } \{N = r\} = e^{-h} \frac{h^r}{r!},$$

$$\text{where } h = \int_{-\infty}^{+\infty} P(t) dt.$$

¹ For a different approach to Poisson's distribution see W. FELLER, *Math. Ann.* 113 (1937) in particular pp. 113-160.

Our proof is based on considerations of characteristic functions. Let $\psi_{\Delta}(x)$ be 1 if x belongs to the interval Δ and 0 otherwise. Thus

$$M(\Delta; N) = \sum_{k=1}^N \psi_{\Delta}(t_k).$$

From the independence of $M(\Delta_1; N)$ and $M(\Delta_2; N)$ it follows that for every pair of real numbers ξ and η we have²

$$\begin{aligned} E \left[\exp \left\{ i \left(\xi \sum_{k=1}^N \psi_{\Delta_1}(t_k) + \eta \sum_{k=1}^N \psi_{\Delta_2}(t_k) \right) \right\} \right] \\ = E \left[\exp \left\{ i \xi \sum_{k=1}^N \psi_{\Delta_1}(t_k) \right\} \right] E \left[\exp \left\{ i \eta \sum_{k=1}^N \psi_{\Delta_2}(t_k) \right\} \right], \end{aligned}$$

where $E[x]$ denotes the mathematical expectation, or mean value, of x . Letting $q(r) = \text{Prob} \{N = r\}$ and using first the independence of N and the t_k 's and then the fact that the t_k 's are independent and identically distributed we obtain

$$\begin{aligned} (2) \quad \sum_{r=0}^{\infty} q(r) (E[\exp \{i(\xi \psi_{\Delta_1}(t) + \eta \psi_{\Delta_2}(t))\}])^r \\ = \sum_{r=0}^{\infty} q(r) (E[\exp \{i \xi \psi_{\Delta_1}(t)\}])^r \sum_{r=0}^{\infty} q(r) (E[\exp \{i \eta \psi_{\Delta_2}(t)\}])^r. \end{aligned}$$

An easy calculation gives

$$E[\exp \{i \xi \psi_{\Delta_1}(t)\}] = 1 + (e^{i\xi} - 1) \int_{\Delta_1} p(t) dt,$$

$$E[\exp \{i \eta \psi_{\Delta_2}(t)\}] = 1 + (e^{i\eta} - 1) \int_{\Delta_2} p(t) dt,$$

$$E[\exp \{i(\xi \psi_{\Delta_1}(t) + \eta \psi_{\Delta_2}(t))\}] = 1 + (e^{i\xi} - 1) \int_{\Delta_1} p(t) dt + (e^{i\eta} - 1) \int_{\Delta_2} p(t) dt.$$

The last equation follows from the fact that Δ_1 and Δ_2 do not overlap. Putting $\xi = \eta = \pi$, $x = 1 - 2 \int_{\Delta_1} p(t) dt$, $y = 1 - 2 \int_{\Delta_2} p(t) dt$, $\varphi(x) = \sum q(r) x^r$ we see that (2) yields the functional equation

$$(3) \quad \varphi(x + y - 1) = \varphi(x)\varphi(y).$$

One cannot ascertain that (3) holds for all real x and y . First of all the defining power series of $\varphi(x)$ is not known to converge outside the unit circle and secondly it is not obvious that each pair of real numbers x, y between -1 and 1 is such that non-overlapping intervals Δ_1, Δ_2 exist for which

$$x = 1 - 2 \int_{\Delta_1} p(t) dt \quad \text{and} \quad y = 1 - 2 \int_{\Delta_2} p(t) dt.$$

² We use the symbol \bar{R} and $E[R]$ interchangeably to denote the average (mathematical expectation) of R .

However, if one restricts oneself to small Δ_1 and Δ_2 the functional equation (3) is seen to hold in a sufficiently small neighborhood of 1. This is sufficient (in view of the analyticity of φ in the unit circle) to determine $\varphi(x)$.

In fact, differentiating (3) first with respect to x and then with respect to y we get

$$\varphi'(x)\varphi'(y) = \varphi''(x + y - 1).$$

Letting $y = 1$ and putting $\varphi'(1) = h$ we have

$$\varphi''(x) = h\varphi'(x),$$

which yields immediately

$$\varphi(x) = Ae^{hx} + B.$$

An entirely elementary reasoning (which employs the fact that $Ae^{hx} + B$ must satisfy (3)) leads to the conclusion that $B = 0$, $A = e^{-h}$ which in turn implies at once that

$$q(r) = e^{-h} \frac{h^r}{r!}.$$

Finally,

$$\begin{aligned} \int_{\Delta} P(t) dt &= E[M(N; \Delta)] = E\left[\sum_{k=1}^N \psi_{\Delta}(t_k)\right] \\ &= \left(\int_{\Delta} p(t) dt\right) e^{-h} \sum_{r=0}^{\infty} \frac{h^{r+1}}{r!} = h \int_{\Delta} p(t) dt, \end{aligned}$$

and therefore

$$\int_{-\infty}^{+\infty} P(t) dt = h, \quad P(t) = hp(t).$$

Since h is the mean value of N (i.e. \bar{N}) we shall use \bar{N} instead of h .

3. Fourier coefficients of $F(t)$ and their statistical properties. In physical applications it is often convenient to assume that the "pulse function" $f(t)$ is periodic with period T (T large) and one might therefore restrict oneself to the interval $(0, T)$.

It is furthermore assumed that both $f(t)$ and $P(t)$ are sufficiently smooth³ so as to justify the formal operations on Fourier series performed below. Since we work in the interval $(0, T)$ we assume that $P(t) = 0$ for $t < 0$ and $t > T$.

Expanding $f(t)$ in a Fourier series in $(0, T)$ we get

$$f(t) \sim \sum_{-\infty}^{\infty} a(\omega_k) \exp(i\omega_k t), \quad \omega_k = \frac{2\pi k}{T},$$

³ For instance $f(t)$ and $P(t)$ may be assumed to be of bounded variation. Actually, much less severe restrictions suffice but in investigations of this sort far reaching generality would only impair the exposition.

and thus

$$F(t) \sim \sum_{-\infty}^{\infty} a(\omega_k) b(\omega_k) \exp(i\omega_k t),$$

where

$$b(\omega_k) = \sum_{j=1}^N \exp(-i\omega_k t_j).$$

Note that

$$\begin{aligned} E[\exp(-i\omega t)] &= \int_0^T \exp(-i\omega t) p(t) dt \\ &= \frac{1}{T} \int_0^T \exp(-i\omega t) P(t) dt \\ &= \frac{1}{T} \int_0^T P(t) dt \quad \frac{\rho(\omega)}{\rho(0)} = c(\omega) + i s(\omega) \end{aligned}$$

and put

$$\begin{aligned} E[F(t)] &= \frac{\tilde{N}}{\rho(0)} \sum_{-\infty}^{+\infty} a(\omega_k) \rho(\omega_k) \exp(i\omega_k t) = T \sum_{-\infty}^{+\infty} a(\omega_k) \rho(\omega_k) \exp(i\omega_k t), \\ X_k^{(\tilde{N})} &= \frac{\sum_{j=1}^N \cos(\omega_k t_j) - \tilde{N} c(\omega_k)}{\sqrt{\tilde{N}}}, \\ Y_k^{(\tilde{N})} &= \frac{\sum_{j=1}^N \sin(\omega_k t_j) - \tilde{N} s(\omega_k)}{\sqrt{\tilde{N}}}. \end{aligned}$$

Thus remembering that $\tilde{N} = \int_0^T P(t) dt$ we may write

$$\frac{F(t) - \overline{F(t)}}{\sqrt{\tilde{N}}} \sim \sum_{-\infty}^{+\infty} a(\omega_k) (X_k^{(\tilde{N})} - iY_k^{(\tilde{N})}) \exp(i\omega_k t)$$

or

$$\frac{F(t) - \overline{F(t)}}{\sqrt{\rho(0)}} \sim \sqrt{T} \sum_{-\infty}^{\infty} a(\omega_k) (X_k^{(\tilde{N})} - iY_k^{(\tilde{N})}) \exp(i\omega_k t).$$

We can now state the following:

THEOREM 2. *In the limit as $\tilde{N} \rightarrow \infty$ each $X_k^{(\tilde{N})}$ (and $Y_k^{(\tilde{N})}$) is normally distributed with mean 0 and variance $\frac{1}{2} + \frac{1}{2}c(2\omega_k)$ ($\frac{1}{2} - \frac{1}{2}c(2\omega_k)$).*

The proof, as usual, is based on the consideration of the characteristic function of $X_k^{(\tilde{N})}$.

We have

$$\begin{aligned} E[\exp \{i\xi X_k^{(\bar{N})}\}] \\ &= \exp \{-i\xi\sqrt{\bar{N}} c(\omega_k)\} \exp(-\bar{N}) \sum_{r=0}^{\infty} \frac{(\bar{N})^r}{r!} \left(E \left[\exp \left\{ \frac{i\xi \cos \omega_k l}{\sqrt{\bar{N}}} \right\} \right] \right)^r \\ &= \exp \{-i\xi\sqrt{\bar{N}} c(\omega_k)\} \exp(-\bar{N}) \exp \left\{ \bar{N} E \left[\exp \left\{ \frac{i\xi \cos \omega_k l}{\sqrt{\bar{N}}} \right\} \right] \right\}. \end{aligned}$$

In deriving this formula use has been made of the facts that the l 's are independent and identically distributed, that N is independent of the l 's and that N is distributed according to Poisson's law. It is now easy to see that as $\bar{N} \rightarrow \infty$ the characteristic function of $X_k^{(\bar{N})}$ approaches

$$\exp \left\{ -\left(\frac{1}{4} + \frac{1}{4}c(2\omega_k) \right) \xi^2 \right\}$$

uniformly in every finite ξ -interval. This, in view of the continuity theorem for Fourier-Stieljes transforms, implies our theorem. It should be mentioned that it is tacitly assumed that even though $\bar{N} = T\rho(0)$ approaches ∞ it does it in such a way that the ratio $\rho(\omega)/\rho(0)$ (and hence $c(\omega)$) remains constant (or more generally, approaches a limit).

By considering the characteristic function of the joint distribution of $X_k^{(\bar{N})}$ and $X_l^{(\bar{N})}$ ($|k| \neq |l|$) (or any other pair like, for instance, $X_k^{(\bar{N})}$ and $Y_l^{(\bar{N})}$, in which case no restriction on k, l is necessary) we are able to prove

THEOREM 3. *In the limit as $\bar{N} \rightarrow \infty$ the distinct Fourier coefficients of $(F(t) - E[F(t)])/\sqrt{\rho(0)}$ are normally correlated (i.e. their joint distribution function is the bivariate normal distribution).*

It is also clear that the higher correlations (i.e. between more than two coefficients) will lead to multivariate normal distributions with coefficients expressible in terms of Fourier coefficients of $P(t)$.

We do not state Theorem 3 in more definite terms because in the next section we shall give a more convenient and useful way of handling correlation properties of our Fourier coefficients.

4. Statistical structure of Fourier coefficients. Let us assume that $P(t) > \gamma > 0$ and that the Fourier series of $P(t)$ converges everywhere.

Expanding $\sqrt{P(t)}$ in a Fourier series in $(0, T)$ we have

$$\sqrt{P(t)} = \sum_{-\infty}^{\infty} \sigma(\omega_l) \exp(i\omega_l t),$$

and in particular (since $p(t) = P(t)/\bar{N}$)

$$\sqrt{p(t)} = \frac{1}{\sqrt{\bar{N}}} \sum_{-\infty}^{\infty} \sigma(\omega_l) \exp(i\omega_l t).$$

We can now write

$$\begin{aligned}
 b(\omega_k) &= \sum_{j=1}^N \exp(-i\omega_k t_j) = \sum_{j=1}^N \frac{\exp(-i\omega_k t_j)}{\sqrt{p(t_j)}} \sqrt{p(t_j)} \\
 &= \frac{1}{\sqrt{N}} \sum_{l=-\infty}^{\infty} \sigma(\omega_l) \left\{ \sum_{j=1}^N \frac{\exp(i(\omega_l - \omega_k)t_j)}{\sqrt{p(t_j)}} \right\} \\
 &= \frac{1}{\sqrt{N}} \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) \left\{ \sum_{j=1}^N \frac{\exp(i\omega_l t_j)}{\sqrt{p(t_j)}} \right\} \\
 &= T \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) \sigma(-\omega_l) + \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) \left\{ \frac{1}{\sqrt{N}} \sum_{j=1}^N \frac{\exp(i\omega_l t_j)}{\sqrt{p(t_j)}} - T \sigma(-\omega_l) \right\} \\
 &= T \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) \sigma(-\omega_l) \\
 &\quad + \sqrt{T} \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) \left\{ \frac{1}{\sqrt{N}T} \sum_{j=1}^N \frac{\exp(i\omega_l t_j)}{\sqrt{p(t_j)}} - \sqrt{T} \sigma(-\omega_l) \right\}.
 \end{aligned}$$

Put $\sigma(\omega) = \alpha(\omega) + i\beta(\omega)$, note that by Parseval's relation

$$\rho(\omega_k) = \sum_{l=-\infty}^{\infty} \sigma(\omega_k + \omega_l) \sigma(-\omega_l),$$

and introduce random variables $U_l^{(N)}$ and $V_l^{(N)}$ by means of the formulas

$$\begin{aligned}
 U_l^{(N)} &= \frac{1}{\sqrt{N}T} \sum_{j=1}^N \frac{\cos(\omega_l t_j)}{\sqrt{p(t_j)}} - \sqrt{T} \alpha(\omega_l), \\
 V_l^{(N)} &= \frac{1}{\sqrt{N}T} \sum_{j=1}^N \frac{\sin(\omega_l t_j)}{\sqrt{p(t_j)}} - \sqrt{T} \beta(\omega_l).
 \end{aligned}$$

Thus

$$b(\omega_k) = T\rho(\omega_k) + \sqrt{T} \sum_{l=-\infty}^{\infty} \sigma(\omega_k + \omega_l) (U_l^{(N)} - iV_l^{(N)})$$

and we have the following theorem.

THEOREM 4. *In the limit as $N \rightarrow \infty$ the random variables $U_0^{(N)}$, $U_1^{(N)}$, $V_1^{(N)}$, $U_2^{(N)}$, $V_2^{(N)}$, \dots are independent and normally distributed (each with mean 0 and variance $\frac{1}{2}$).*

This theorem can be proved in a manner exactly analogous to that of Theorem 2. We need only consider the characteristic functions of the joint distributions of U 's and V 's and treat them in the same way as we treated the characteristic function of the distribution of a single X in the proof of Theorem 2. One thing, however, should be strongly emphasized. The proof of independence (in the limit as $N \rightarrow \infty$) of $U_l^{(N)}$ and $U_m^{(N)}$ ($|l| \neq |m|$), for instance, depends on proving that

$$E[U_l^{(N)} U_m^{(N)}] = 0.$$

This in turn depends essentially on the fact that N is distributed according to Poisson's law.

In fact,

$$E[U_i^{(\bar{N})} U_m^{(\bar{N})}] = \frac{1}{\bar{N}T} E\left[\left(\sum_{j=1}^N \frac{\cos(\omega_l t_j)}{\sqrt{p(t_j)}}\right)\left(\sum_{j=1}^N \frac{\cos(\omega_m t_j)}{\sqrt{p(t_j)}}\right)\right] = T\alpha(\omega_l)\alpha(\omega_m).$$

But

$$\begin{aligned} E\left[\left(\sum_{j=1}^N \frac{\cos(\omega_l t_j)}{\sqrt{p(t_j)}}\right)\left(\sum_{j=1}^N \frac{\cos(\omega_m t_j)}{\sqrt{p(t_j)}}\right)\right] \\ = E\left[\sum_{j=1}^N \frac{\cos \omega_l t_j \cos \omega_m t_j}{p(t_j)}\right] + E\left[\sum_{j \neq k} \frac{\cos \omega_l t_j \cos \omega_m t_k}{\sqrt{p(t_j)}\sqrt{p(t_k)}}\right] \\ = E\left[\frac{N(N-1)}{(\bar{N})^2}\right] T\alpha(\omega_l)\alpha(\omega_m), \end{aligned}$$

and finally

$$E[U_i^{(\bar{N})} U_m^{(\bar{N})}] = \left(\frac{\bar{N}^2 - \bar{N}}{(\bar{N})^2} - 1\right) T\alpha(\omega_l)\alpha(\omega_m)$$

Since for Poisson's distribution $\bar{N}^2 = \bar{N} + (\bar{N})^2$ we get

$$E[U_i^{(\bar{N})} U_m^{(\bar{N})}] = 0.$$

Also the proof that $E[|U_i^{(\bar{N})}|^2] = \frac{1}{2}$ employs essentially the fact that N is distributed according to Poisson's law.

In view of Theorem 4 we can restate Theorem 3 in a form which is both useful and illuminating inasmuch as it describes completely the statistical structure of the $b(\omega_k)$'s and hence of the Fourier coefficients of $F(t)$.

THEOREM 5. *For the purposes of finding correlations between the $b(\omega_k)$'s it suffices to replace each $b(\omega_k)$ (in the limit as $\bar{N} \rightarrow \infty$) by its "statistical representation"*

$$Tp(\omega_k) + \sqrt{T} \sum_{l=-\infty}^{\infty} \sigma(\omega_k + \omega_l) A_l,$$

where A_{-l} is the complex conjugate of A_l , A_0, A_1, A_2, \dots a sequence of independent complex-valued random variables and each A_k is distributed in such a way that $\theta_k = \arg A_k$ is uniformly distributed independent of A_k and the density of the probability distribution of $|A_k|$ is

$$2Ae^{-A^2}, \quad (A \geq 0).$$

Theorem 5 was proved under the assumption $P(t) > \gamma > 0$. This assumption was needed to validate the convenient artifice of multiplying and dividing by $\sqrt{p(t_j)}$.

However, even in the case when $P(t)$ is not bounded from below by a positive

number (it is always true that $P(t) \geq 0$). Theorem 5 remains true. It could be proved by direct but tedious considerations suggested in section 2.

Theorems 4 and 5 can be easily extended to the case when the pulses all have the same shape but may, at random, differ in magnitude. In other words, instead of sum (1) we may consider the sum

$$(4) \quad F(t) = \sum_{j=1}^N \epsilon_j f(t - t_j),$$

where the individual pulses are independent and a function $P(\epsilon, t)$ is given such that

$$\int_{\epsilon}^{\epsilon+\Delta\epsilon} \int_t^{t+\Delta t} P(\epsilon, t) dt d\epsilon$$

is the average number of pulses of "amplitude" between ϵ and $\epsilon + \Delta\epsilon$ occurring between t and $t + \Delta t$.

Theorems 4 and 5 still hold provided one replaces the Fourier coefficients of $P(t)$ by those of

$$\int_{-\infty}^{+\infty} \epsilon P(\epsilon, t) d\epsilon,$$

and the Fourier coefficients of $\sqrt{P(t)}$ by those of

$$\sqrt{Q(t)} = \sqrt{\int_{-\infty}^{+\infty} \epsilon^2 P(\epsilon, t) d\epsilon}.$$

5. Concluding remarks and summary. If one assumes that the number of pulses N in the time interval $(0, T)$ is constant instead of being a random variable obeying Poisson's law, then Theorems 4 and 5 fail. The failure is due to the fact that, for instance $E[U_i^{(\hat{N})} U_m^{(\hat{N})}]$ is no longer 0. However, as $T \rightarrow \infty$ the changes in correlation due to assuming N constant become negligible. On the other hand if one assumes that the number of pulses in each of the time intervals $(0, \tau)$, $(\tau, 2\tau)$, \dots is fixed, the changes in correlations become appreciable. This case can also be treated by the above methods.

The case in which $p(t)$ is independent of time has been considered in various connections by Schottky, Uhlenbeck and Goudsmidt and Rice⁴. Their investigations emphasized the importance and usefulness of the harmonic analysis of random functions.

In conclusion we summarize our results for the case of time-dependent $P(\epsilon, t)$

⁴ W. SCHOTTKY, *Ann. d. Phys.* 57 (1919) pp. 541-587.

G. E. UHLENBECK and S. GOUDSMIDT, *Phys. Rev.* 34 (1920) pp. 145-151.

S. O. RICE, mimeographed notes on mathematical analysis of random noise, as yet unpublished.

The authors are indebted to Mr. Rice for making his notes available to them.

by observing that in applications one may replace $F(t)$ by its "statistical representation"

$$(5) \quad E[F(t)] + \sqrt{T} \sum_{k=-\infty}^{\infty} a(\omega_k) \left\{ \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) A_l \right\} \exp(i\omega_k t),$$

where

$$\omega_k = \frac{2\pi k}{T},$$

$$E[F(t)] = T \sum_{-\infty}^{\infty} a(\omega_k) \rho(\omega_k) \exp(i\omega_k t),$$

$$\int_{-\infty}^{\infty} \epsilon P(\epsilon, t) d\epsilon = \sum_{k=-\infty}^{\infty} \rho(\omega_k) \exp(i\omega_k t),$$

$$\sqrt{Q(t)} = \sqrt{\int_{-\infty}^{\infty} \epsilon^2 P(\epsilon, t) d\epsilon} = \sum_{k=-\infty}^{\infty} \sigma(\omega_k) \exp(i\omega_k t),$$

and the A_l 's are normally distributed complex-valued random variables for which

$$E[A_l] = 0, \quad E[|A_l|^2] = 1, \quad A_l^* = A_{-l}$$

Furthermore, for $l \geq 0$ the A_l 's are statistically independent.

Thus

$$F(t) - E[F(t)] \sim \sum_{k=-\infty}^{\infty} \lambda(\omega_k) \exp(i\omega_k t)$$

where the λ 's are normally distributed complex-valued random variables obeying the relation

$$E[|\lambda(\omega)|^2] = |a(\omega)|^2 \int_0^T Q(t) dt.$$

If $Q(t)$ is periodic with frequency $\frac{\omega_{k_0}}{2\pi}$ then it follows that $\lambda(\omega')$ and $\lambda(\omega'')$ are independent unless $\omega' + \omega''$ or $\omega' - \omega''$ is an integral multiple of ω_{k_0} .

Finally, we mention that $F(t) - E[F(t)]$ is normally distributed with variance $s(t)$ given by the formula

$$s^2(t) = E[(F(t) - E[F(t)])^2] = T \sum_{k=-\infty}^{\infty} \gamma(\omega_k) \mu(\omega_k) \exp(i\omega_k t),$$

where $\gamma(\omega_k)$ is the Fourier coefficient of $Q(t)$ and $\mu(\omega_k)$ the Fourier coefficient of $f^2(t)$.

RANDOM ALMS

BY PAUL R. HALMOS

Syracuse University

1. Statement of the problem. Consider the problem of distributing one pound of gold dust at random among a countably infinite set of beggars. Let the beggars be enumerated and let the procedure for distribution be as follows: the first beggar is given a random portion of the gold; the second beggar gets a random portion of the remainder; \cdots and so on ad infinitum. In this description the phrase "random portion" occurs an infinite number of times: it seems reasonable to require that it have the same interpretation each time. To be precise: let x_j ($j = 0, 1, 2, \cdots$) be the amount received by the j th beggar. Let the distribution of x_0 be given by a density function $p(\lambda)$:

$$(1) \quad p(\lambda) \geq 0, \quad 0 \leq \lambda \leq 1;$$

$$(2) \quad \int_0^1 p(\lambda) d\lambda = 1;$$

$$(3) \quad P(a < x_0 < b) = \int_a^b p(\lambda) d\lambda, \quad 0 \leq a < b \leq 1$$

After the first beggar has received his alms and the amount of gold dust left is μ , (i.e. $x_0 = 1 - \mu$), the value of x_1 will be between 0 and μ . The uniformity requirement mentioned above means that the proportion of μ that the second beggar is to receive is again determined by the probability density p ; in other words the conditional probability that x_1 be between $\lambda\mu$ and $(\lambda + d\lambda)\mu$, given that $x_0 = 1 - \mu$, is $p(\lambda) d\lambda$. In symbols:

$$(4) \quad P(a\mu < x_1 < b\mu | x_0 = 1 - \mu) = \int_a^b p(\lambda) d\lambda.$$

Writing $\alpha = a\mu, \beta = b\mu$, (4) becomes

$$(5) \quad P(\alpha < x_1 < \beta | x_0 = 1 - \mu) = \int_{\alpha/\mu}^{\beta/\mu} \frac{1}{\mu} p\left(\frac{\lambda}{\mu}\right) d\lambda.$$

More generally I shall assume that the conditional probability distribution of x_n , assuming that after the preceding donations there is left an amount μ , is given in the interval $(0, \mu)$ by $\frac{1}{\mu} p\left(\frac{\lambda}{\mu}\right)$. In symbols:

$$(6) \quad P(a < x_n < b | \sum_{i < n} x_i = 1 - \mu) = \int_a^b \frac{1}{\mu} p\left(\frac{\lambda}{\mu}\right) d\lambda, \quad 0 \leq a < b \leq \mu.$$

This assumption completely determines (in terms of p) the joint distribution of the whole infinite sequence $\{x_0, x_1, x_2, \cdots\}$. Several interesting special

questions may be asked about this distribution. For example: What are the expectation, dispersion, and higher moments of the x_n ? What, similarly, are the moments of the partial sum $S_n = \sum_{j \leq n} x_j$? More generally what are the exact distributions of x_n and of S_n ? Will the process described really distribute all the gold, or is there a positive probability that some is left even after every beggar had his turn? What is the rate of convergence of the series $\sum_{n \geq 0} x_n$? It is the purpose of this paper to answer these and a few related questions.

2. Calculation of distributions. The $n + 1$ dimensional probability density of the distribution of (x_0, x_1, \dots, x_n) is given by¹

$$(7) \quad \prod_{i \leq n} \frac{1}{1 - \sum_{j < i} \lambda_j} p \left(\frac{\lambda_i}{1 - \sum_{j < i} \lambda_j} \right)$$

in the region defined by $\lambda_j \geq 0$, $\lambda_0 + \dots + \lambda_n \leq 1$. For $n = 0$ there is only one term in the product and that one is equal to $p(\lambda_0)$, the region is defined by $0 \leq \lambda_0 \leq 1$. The formula reduces in this case to the definition of the distribution of x_0 . The general case follows inductively by the use of the conditional probability formula (6). (For example: $P(x_0 = \lambda_0, x_1 = \lambda_1) = P(x_0 = \lambda_0)P(x_1 = \lambda_1 | x_0 = \lambda_0) = p(\lambda_0) \frac{1}{1 - \lambda_0} p \left(\frac{\lambda_1}{1 - \lambda_0} \right)$.)

From (7) it is possible in principle to calculate the densities of the distributions of x_n and of S_n . Thus for example the density q_n of the distribution of x_n is found by integrating out the λ_j with $j < n$ from (7), so that

$$(8) \quad q_n(\lambda_n) = \int \dots \int \prod_{i \leq n} \frac{1}{1 - \sum_{j < i} \lambda_j} p \left(\frac{\lambda_i}{1 - \sum_{j < i} \lambda_j} \right) d\lambda_0 \dots d\lambda_{n-1},$$

where the integration is extended over the region defined by $\lambda_j \geq 0$ ($0 \leq j \leq n$), $\sum_{j \leq n} \lambda_j \leq 1$. Similarly $V_n(t) = P(S_n < t)$ is given by

$$(9) \quad V_n(t) = \int \dots \int \prod_{i \leq n} \frac{1}{1 - \sum_{j < i} \lambda_j} p \left(\frac{\lambda_i}{1 - \sum_{j < i} \lambda_j} \right) d\lambda_0 \dots d\lambda_n,$$

($0 \leq t \leq 1$) where the domain of integration is defined by $\lambda_j \geq 0$ ($0 \leq j \leq n$), $\sum_{j \leq n} \lambda_j < t$.

Working with integrals of the type (8) and (9) is often greatly facilitated by the substitution $\mu_i = \sum_{j \leq i} \lambda_j$, ($\lambda_i = \mu_i - \mu_{i-1}$), $0 \leq i \leq n$. The Jacobian of this linear change of variables is identically one. The domain of integration used in (9) is defined in terms of the μ 's by $0 \leq \mu_0 \leq \mu_1 \leq \dots \leq \mu_n \leq t \leq 1$, so that

$$(10) \quad V_n(t) = \int_0^t d\mu_n \int_0^{\mu_n} d\mu_{n-1} \dots \int_0^{\mu_1} d\mu_0 \prod_{i \leq n} \frac{1}{1 - \mu_{i-1}} p \left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}} \right).$$

¹ A summation or a product extended over an empty set of indices will, as is customary, be interpreted as 0 or 1 respectively. Since throughout this paper only non-negative indices are considered, whenever the notation indicates a negative index the quantity to which it is attached is to be interpreted as 0.

Hence the density of the distribution of S_n is

$$(11) \quad v_n(t) = \int_0^t d\mu_{n-1} \int_0^{\mu_{n-1}} d\mu_{n-2} \cdots \int_0^{\mu_1} d\mu_0 \prod_{i=1}^n \frac{1}{1 - \mu_{i-1}} p\left(\frac{\mu_1 - \mu_{i-1}}{1 - \mu_{i-1}}\right) \\ \cdot \frac{1}{1 - \mu_{n-1}} p\left(\frac{t - \mu_{n-1}}{1 - \mu_{n-1}}\right).$$

For later purposes it is more convenient to set $t = \mu_n$ in (11) and to express $v_n(\mu_n)$ as a multiple (and not as an iterated) integral; then

$$(12) \quad v_n(\mu_n) = \int \cdots \int \prod_{i=1}^n \frac{1}{1 - \mu_{i-1}} p\left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}}\right) d\mu_0 \cdots d\mu_{n-1},$$

where the domain of integration is defined by $0 \leq \mu_0 \leq \mu_1 \leq \cdots \leq \mu_{n-1} \leq \mu_n \leq 1$. The integrals (8) and (12) are explicitly evaluated below for a special case.

It is possible from (8) to find the k th moment $M_k^{(n)}$ of x_n , $M_k^{(n)} = \int_0^1 \lambda_n^k q_n(\lambda_n) d\lambda_n$. Write

$$\alpha_k = \int_0^1 \lambda^k p(\lambda) d\lambda, \quad \beta_k = \int_0^1 (1 - \lambda)^k p(\lambda) d\lambda.$$

Clearly $M_k^{(n)}$ is obtained from (8) upon multiplication by λ_n^k and integration with respect to λ_n .

$$(13) \quad M_k^{(n)} = \int \cdots \int \lambda_n^k \prod_{i=1}^n \frac{1}{1 - \sum_{j < i} \lambda_j} p\left(1 - \sum_{j < i} \lambda_j\right) d\lambda_0 \cdots d\lambda_n.$$

It is advantageous once again to write $\mu_i = \sum_{j \leq i} \lambda_j$. The resulting integral may be written in the iterated form as follows:

$$(14) \quad M_k^{(n)} = \int_0^1 d\mu_0 \int_{\mu_0}^1 d\mu_1 \cdots \int_{\mu_{n-1}}^1 d\mu_n \prod_{i=1}^n \frac{1}{1 - \mu_{i-1}} \\ \cdot p\left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}}\right) \cdot (\mu_n - \mu_{n-1})^k.$$

Consider separately the innermost integral

$$J = \int_{\mu_{n-1}}^1 p\left(\frac{\mu_n - \mu_{n-1}}{1 - \mu_{n-1}}\right) (\mu_n - \mu_{n-1})^k \frac{d\mu_n}{1 - \mu_{n-1}}.$$

Writing $\lambda = (\mu_n - \mu_{n-1})/(1 - \mu_{n-1})$ this becomes

$$J = \int_0^1 p(\lambda) \lambda^k (1 - \mu_{n-1})^k d\lambda = \alpha_k (1 - \mu_{n-1})^k.$$

Hence

$$(15) \quad M_k^{(n)} = \alpha_k \int_0^1 d\mu_0 \cdots \int_{\mu_{n-2}}^1 d\mu_{n-1} \prod_{i=1}^n \frac{1}{1 - \mu_{i-1}} p\left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}}\right) \cdot (1 - \mu_{n-1})^k.$$

The innermost integral this time is

$$J' = \int_{\mu_{n-2}}^1 p \left(\frac{\mu_{n-1} - \mu_{n-2}}{1 - \mu_{n-2}} \right) (1 - \mu_{n-1})^k \frac{d\mu_{n-1}}{1 - \mu_{n-2}}.$$

Write $\lambda = (\mu_{n-1} - \mu_{n-2})/(1 - \mu_{n-2})$; then $(1 - \mu_{n-1}) = (1 - \lambda)(1 - \mu_{n-2})$ and

$$J' = \int_0^1 p(\lambda)(1 - \lambda)^k (1 - \mu_{n-2})^k d\lambda = \beta_k (1 - \mu_{n-2})^k.$$

Hence, finally,

$$(16) \quad M_k^{(n)} = \alpha_k \beta_k \int_0^1 d\mu_0 \cdots \int_{\mu_{n-2}}^1 d\mu_{n-1} \prod_{i < n-1} \frac{1}{1 - \mu_{i-1}} p \left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}} \right) \cdot (1 - \mu_{n-2})^k.$$

Observe now that the right member of (16) (except for the factor β_k) may be obtained from (15) upon replacing n by $n - 1$. In other words $M_k^{(n)} = \beta_k M_k^{(n-1)}$. Since $M_k^{(0)} = \alpha_k$, it follows that

$$(17) \quad M_i^{(n)} = \alpha_k \beta_k^n, \quad n = 0, 1, 2, \dots$$

Instead of calculating similarly the moments $\int_0^1 \mu_n^k v_n(\mu_n) d\mu_n$ of S_n it is more convenient to calculate the quantities

$$N_k^{(n)} = \int_0^1 (1 - \mu_n)^k v_n(\mu_n) d\mu_n.$$

The moments themselves may be obtained from the N 's by simple combinatorial formulas.

It follows from (12) that

$$(18) \quad N_k^{(n)} = \int_0^1 d\mu_0 \int_{\mu_0}^1 d\mu_1 \cdots \int_{\mu_{n-1}}^1 d\mu_n \prod_{i \leq n} \frac{1}{1 - \mu_{i-1}} p \left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}} \right) (1 - \mu_n)^k.$$

The innermost integral in (18) is

$$J'' = \int_{\mu_{n-1}}^1 p \left(\frac{\mu_n - \mu_{n-1}}{1 - \mu_{n-1}} \right) (1 - \mu_n)^k \frac{d\mu_n}{1 - \mu_{n-1}}.$$

Writing $\lambda = (\mu_n - \mu_{n-1})/(1 - \mu_{n-1})$, $(1 - \mu_n)$ becomes $(1 - \lambda)(1 - \mu_{n-1})$, so that

$$J'' = \int_0^1 p(\lambda)(1 - \lambda)^k (1 - \mu_{n-1})^k d\lambda = \beta_k (1 - \mu_{n-1})^k.$$

Consequently

$$(19) \quad \begin{aligned} N_k^{(n)} &= \beta_k \int_0^1 d\mu_0 \cdots \int_{\mu_{n-2}}^1 d\mu_{n-1} \prod_{i < n} \frac{1}{1 - \mu_{i-1}} p \left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}} \right) \cdot (1 - \mu_{n-1})^k \\ &= \beta_k N_k^{(n-1)}, \end{aligned}$$

so that

$$(20) \quad N_k^{(n)} = \beta_k^{n+1}, \quad n = 0, 1, 2, \dots,$$

The additivity of the first moment yields an amusing check on (17) and (20). Since $E(S_n) = E(\sum_{j \leq n} x_j) = \sum_{j \leq n} E(x_j)$ (where E denotes expectation, or first moment), it should be true that $1 = N_1^{(n)} = \sum_{j \leq n} M_1^{(j)}$. In terms of α 's and β 's this means $1 = \beta_1^{n+1} = \alpha_1 \sum_{j \leq n} \beta_1^j$, and this in turn reduces to the trivial identity $\alpha_1 = 1 - \beta_1$.

Since $0 \leq \sum_{j \leq n} x_j \leq 1$ with probability 1 for every n , it is clear that the series $\sum_{j \geq 0} x_j$ converges with probability 1 to a sum x , $0 \leq x \leq 1$. Since $E(x_j) = \alpha_j \beta_1^j$ and since $E(x) = \sum_{j \geq 0} E(x_j)$, it follows that $E(x) = \sum_{j \geq 0} \alpha_j \beta_1^j = \alpha_1 / (1 - \beta_1) = 1$. This implies (since $0 \leq x \leq 1$) that x must be equal to 1 with probability 1. In other words it is almost certain that all the gold dust will eventually be distributed.

3. Product representation. Considerable light is shed on some of the above computations (and in fact the moment formulas (17) and (20) are proved anew) by the following considerations. The principle of equitable treatment enunciated in the introductory paragraph was subsequently formalized by the conditional probability relation (6). It may also be formalized by the following (equivalent) procedure. Let y_0, y_1, y_2, \dots be a sequence of *independent* chance variables each of whose distributions is given by the probability density p ; let y_n be interpreted as the proportion, of the amount available to the n th beggar, that he actually receives. In other words

$$(21) \quad x_n = y_n(1 - \sum_{j < n} x_j), \quad n = 0, 1, 2, \dots$$

The first main problem in this formulation is to express the x 's in terms of the y 's. This is most easily accomplished by an inductive proof of the formula

$$(22) \quad \sum_{i \leq n} x_i = 1 - \prod_{i \leq n} (1 - y_i).$$

For $n = 0$, (22) asserts merely that $x_0 = y_0$. The inductive step proceeds as follows:

$$\begin{aligned} \sum_{i \leq n} x_i &= x_n + \sum_{i \leq n-1} x_i = y_n(1 - \sum_{i \leq n-1} x_i) + \sum_{i \leq n-1} x_i \\ &= y_n \prod_{i \leq n-1} (1 - y_i) + 1 - \prod_{i \leq n-1} (1 - y_i) \\ &= 1 - (1 - y_n) \prod_{i \leq n-1} (1 - y_i) = 1 - \prod_{i \leq n} (1 - y_i). \end{aligned}$$

From (22) it follows that

$$(23) \quad x_n = y_n \prod_{i < n} (1 - y_i)$$

and

$$(24) \quad R_n = 1 - S_n = 1 - \sum_{j \leq n} x_j = \prod_{j \leq n} (1 - y_j).$$

The moment formulas (17) and (20) follow immediately from (23) and (24) respectively

Another very important application of (23) and (24) is the following theorem. If the first geometric moment (geometric mean)

$$r = \exp \{E(\log [1 - y_1])\} = \exp \left\{ \int_0^1 \log (1 - \lambda) p(\lambda) d\lambda \right\}$$

is different from zero (i.e. if $\int_0^1 \log (1 - \lambda) p(\lambda) d\lambda$ is finite) then the limits

$$\lim_{n \rightarrow \infty} (x_n/y_n)^{1/n} \quad \text{and} \quad \lim_{n \rightarrow \infty} R_n^{1/n}$$

both exist and are both equal to r .

Since according to (23) and (24), $x_n/y_n = R_{n-1}$ the two parts of the conclusion are seen to be equivalent. For the proof take the logarithm of both sides of (24) and divide by n , obtaining

$$(25) \quad \log R_n^{1/n} = \frac{1}{n} \sum_{j \leq n} \log (1 - y_j).$$

Since, according to the hypotheses stated, the chance variables $\log (1 - y_j)$ are independent and all have the same distribution with a finite expectation, the strong law of large numbers applies to the right side of (25) and (after taking exponentials) yields the desired conclusion.

The result just obtained may be phrased as follows: with probability 1 x_n is asymptotically equal to $r^n y_n$. This statement shows that in an obvious if somewhat crude sense the rate of convergence of $\sum_{j \geq 0} x_j$ is that (at least) of a geometric series with ratio r . This conclusion is further supported by the behavior of R_n , which again is the sort of thing one expects from a geometric series. (That is: the n th root of the n th remainder of a geometric series always does converge to the common ratio) As usual, more delicate quantitative results concerning the rate of convergence may be obtained by applying to (25) not merely the law of large numbers but the law of the iterated logarithm.

The product representation of x_n in formula (23) points the way to a generalization of this theory which may be of some interest. In this generalization x_n is still defined by (23) and the y 's are still independent, but the distribution of y_j is given by a density p_j , where the p 's need not be equal to each other. In terms of random alms this means that the condition of equitable treatment is replaced by the following weaker condition: the probability distribution of the amount that the j th beggar receives depends only on j and on the amount left by the preceding beggars, and in particular does not depend on the sizes of the alms already distributed. Many of the conclusions obtained under the simpler

hypotheses carry over to this generalized case with only slight changes. In particular the distribution formulas (7), (8), and (12), and the moment formulas (17) and (20), are changed only to the extent of acquiring an extra subscript due to the difference of the p_j .

4. Applications. (A) The original motivation of the present work was an investigation of the notion of a random mass distribution, and the results obtained may be considered as one possible solution of the problem of defining randomness for mass distributions in the special (discrete) case where the entire mass is concentrated on the non-negative integers. It would be of great interest to extend the results of this note to various continuous cases in which the set of integers is replaced by the unit interval, or the entire real line, or n dimensional Euclidean space. I intend to study some of these extensions at another time; at the moment I merely mention one implication of this statistical point of view.

Considering the sequence $\{x_0, x_1, x_2, \dots\}$ as a system of weights, the integer n carrying the weight x_n , various questions may be raised concerning properties of the discrete mass distributions so obtained. For example: do the moments $m_k = \sum_{n \geq 0} n^k x_n$ exist and, if so, what are their averages and dispersions and, more generally, their moments and their distributions? I shall settle here the questions concerning existence and expectation.

The chance variable m_k is non-negative and, even if it is infinite with positive probability, its expectation is defined by $E(m_k) = \sum_{n \geq 0} n^k E(x_n) = \sum_{n \geq 0} n^k M_1^{(n)} = \sum_{n \geq 0} n^k \alpha_1 \beta_1^n$. Since $0 < \beta_1 < 1$, the last written series converges and therefore $E(m_k)$ is finite. This implies that m_k is finite with probability 1.

(B) It has been observed that the logarithms of the sizes of particles such as mineral grains are frequently normally distributed. Kolmogoroff² has given an explanation of this phenomenon; the results of the present paper yield an alternative and in some respects simpler explanation. Suppose in fact that the probability of a particle losing a chip the proportion of whose size to the size of the original particle is between λ and $\lambda + d\lambda$ is $p(\lambda) d\lambda$. With this stochastic scheme the size of the remaining particle after n chips have been lost is given by R_n . Since, by (25), $\log R_n$ is a sum of independent chance variables with the same distributions, the Laplace-Lapounoff theorem may be invoked to show that the distribution of R_n is for large n nearly normal. (It is necessary of course to assume here the finiteness of the second geometric moment, or equivalently of the integral $\int_0^1 \log^2 (1 - \lambda) p(\lambda) d\lambda$.) The mean and the variance of each summand of $\log R_n$ are

$$a = \int_0^1 \log (1 - \lambda) p(\lambda) d\lambda \quad \text{and} \quad b^2 = \int_0^1 [\log (1 - \lambda) - a]^2 p(\lambda) d\lambda,$$

² A. N. Kolmogoroff, "Ueber das logarithmisch normale Verteilungsgesetz der Dimensionen der Teilchen bei Zerstückelung," *C. R. (Doklady) Acad. Sci. U.R.S.S.* (N. S.) Vol. 31 (1941), pp. 99-101.

respectively; consequently (by the additivity of the mean and the variance) the corresponding parameters of the distribution of $\log R_n$ (and hence of the approximating normal distribution) are given by $(n+1)a$ and $(n+1)b^2$ respectively.

(C) A special case of the distributions studied in this paper (namely the case of uniform distribution, $p(\lambda) = 1$) arises in the theory of scattering of neutrons by protons of the same mass. According to Bethe³. "In each collision with a proton the neutron will lose energy. As long as the neutron is fast compared to the proton, the probability that the neutron energy lies between E and $E + dE$ after the collision, is $w(E) dE = dE/E_0$, where E_0 is the neutron energy before the collision. This means that any value of the final energy of the neutron, between 0 and the initial energy E_0 , is equally probable."

To calculate explicitly the distributions it is most convenient to start from (11). If p (with any argument) is replaced by 1 and the terms of the product are distributed, each under its own differential, (11) takes the form

$$(26) \quad v_n(t) = \int_0^t \frac{d\mu_{n-1}}{1 - \mu_{n-1}} \int_0^{\mu_{n-1}} \frac{d\mu_{n-2}}{1 - \mu_{n-2}} \cdots \int_0^{\mu_1} \frac{d\mu_0}{1 - \mu_0}.$$

The value of the iterated integral is easy to obtain: $v_n(t) = (-1)^n(1/n!) \log^n(1-t)$. Since $v_n(t)$ gives the distribution of the partial sum S_n , the distribution of $R_n = 1 - S_n$ is given by $v_n(1-t) = (-1)^n(1/n!) \log^n t$.⁴ It is possible but not necessary to derive similarly the distribution of x_n . It is simpler to obtain this distribution by exploiting the symmetry of the uniform distribution. Since, according to (23) and (24), x_n and R_n are both products of $n+1$ uniformly and independently distributed chance variables they have the same distribution, so that the density of the distribution of x_n is also given by $(-1)^n(1/n!) \log^n t$, $n = 0, 1, 2, \dots$

The roles of the geometric mean r ($= 1/e$ in case $p = 1$) and of the normal distribution have also been observed in the physical situation. Fermi⁵ has expressed the geometric series like behavior of $\sum_{n \geq 0} x_n$ by the statement "... an impact of a neutron against a proton reduces, on the average, the neutron energy by a factor $1/e$," and Bethe⁶ remarks that "... the actual values of $\log E$ after n collisions form very nearly a Gaussian distribution ..."

³ H. A. Bethe, "Nuclear Physics, B Nuclear Dynamics, Theoretical," *Reviews of Modern Physics*, Vol. 9(1937) p 120

⁴ This distribution has been calculated by E. U. Condon and G. Breit, "The energy distribution of neutrons slowed by elastic impacts," *Physical Review*, Vol. 49(1936) pp. 229-231.

⁵ Quoted by Condon and Breit, *loc. cit.*

⁶ *Loc. cit.*

ON BIASES IN ESTIMATION DUE TO THE USE OF PRELIMINARY TESTS OF SIGNIFICANCE

BY T. A. BANCROFT

Iowa State College

1. INTRODUCTION

In problems of statistical estimation we often express the joint frequency distribution of the sample observations x_1, x_2, \dots, x_n in the form

$$(1) \quad f(x_1, \dots, x_n; \alpha, \beta, \gamma, \dots) \Pi dx_i, \quad (i = 1, \dots, n)$$

where the functional form, f , is assumed known, and $\alpha, \beta, \gamma, \dots$ are certain population parameters whose values may or may not be known. Given this specification, statistical theory provides routine mathematical processes for obtaining estimates of the parameters $\alpha, \beta, \gamma, \dots$ from the observations x_1, x_2, \dots, x_n .

In performing tests of significance we often assume that the data follow some distribution

$$(2) \quad f_1(x_1, \dots, x_n; \alpha, \beta, \gamma, \dots) \Pi dx_i, \quad (i = 1, \dots, n)$$

where f_1 is a known function or family of functions. We may wish to test the hypothesis that the data follow the more specialized distribution

$$(3) \quad f_2(x_1, \dots, x_n; \alpha', \beta', \gamma', \dots) \Pi dx_i, \quad (i = 1, \dots, n)$$

where f_2 is some member or sub-group of the family f_1 . Given this specification, statistical theory provides routine mathematical processes for testing such hypotheses.

In the application of statistical theory to specific data, there is often some uncertainty about the appropriate specifications in equations (1), (2) and (3). In such cases preliminary tests of significance have been used, in practice, as an aid in choosing a specification. We shall give several examples from the literature of statistical methodology.

(1) In an analysis of variance, in order to obtain a best estimate of variance, we may be uncertain as to whether two mean squares in the lines of the analysis may be assumed homogeneous, [1]. Suppose that it is desired to estimate the variance σ_1^2 , of which an unbiased estimate s_1^2 is available. In addition, there is an unbiased estimate s_2^2 of σ_2^2 , where from the nature of the data it is known that either $\sigma_2^2 = \sigma_1^2$ or $\sigma_2^2 < \sigma_1^2$. As a criterion in making a decision the following rule of procedure is used frequently: test s_1^2/s_2^2 by the F -test, where s_1^2 and s_2^2 are the two mean squares. If F is not significant at some assigned significance level use $(n_1 s_1^2 + n_2 s_2^2)/(n_1 + n_2)$ as the estimate of σ_1^2 . If F is significant at the assigned significance level, use s_1^2 as the estimate of σ_1^2 .

(2) After working out the regression of y on a number of independent variates we may be uncertain as to the appropriateness of the retention of some one of

the independent variates, [2]. To illustrate let us consider the choice between the regression equations $y = b_1x_1 + b_2x_2$ and $y' = b'_1x_1$, after having fitted $y = b_1x_1 + b_2x_2$; the population regression equation being $y = \beta_1x_1 + \beta_2x_2$. In this case a procedure commonly used in deciding whether to retain x_2 is as follows: we test s_2^2/s_3^2 by the F -test, where s_2^2 is the reduction in sum of squares due to x_2 after fitting x_1 , and s_3^2 is the residual mean square. If F is not significant at some assigned significance level we omit the term containing x_2 and use b'_1 as the estimate of β_1 . If F is significant we retain the term containing x_2 and use b_1 as the estimate of β_1 . A similar example occurs in fitting a polynomial, when there is uncertainty as to the appropriate degree for the polynomial [3].

(3) In certain analyses we may be uncertain as to the appropriateness of the use of the χ^2 test. Bartlett gives an illustration in a discussion of binomial variation, [4]. He performs two supplementary χ^2 tests of significance as an aid in deciding to abandon the main use of the χ^2 test altogether, and proceeds to use an analysis of variance instead. It is of interest to note that the main use of the χ^2 test gives a significant difference at the 5% level while, in the analysis of variance, Fisher's z is not significant at the 5% level. Here again we might formulate a "rule of procedure" and follow through the analysis as in the preceding cases.

This use of tests of significance as an aid in determining an appropriate specification, and hence the form that the completed analysis shall take, involves acting as if the null hypothesis is false in those cases in which it is refuted at some assigned significance level, and, on the other hand, acting as if the null hypothesis is true in those cases in which we fail to refute it at the assigned significance level. An investigation of the consequences of some of these uses is the purpose of this paper.

It is proposed to consider the first two cases mentioned above: (1) a test of the homogeneity of variances, and (2) a test of a regression coefficient. A complete investigation of the consequences of the rules of procedure would be very extensive, since these consequences depend on the form of the subsequent statistical analysis. As a beginning, it is proposed to limit the study to the efficiency of these "rules of procedure" in the control of bias.

The need for solutions of a whole family of problems of this kind has been pointed out recently by Berkson [5].

II. EXAMPLE ONE: TEST OF HOMOGENEITY OF VARIANCES

1. Statement of the problem. s_1^2 and s_2^2 are two independent estimates of variances σ_1^2 and σ_2^2 respectively, (such that $n_1s_1^2/\sigma_1^2$, $n_2s_2^2/\sigma_2^2$ are distributed independently according to χ_1^2 and χ_2^2 , with n_1 and n_2 degrees of freedom). It is known that $\sigma_2^2 \leq \sigma_1^2$. To obtain from these an estimate of σ_1^2 , to be used in the particular analysis in hand, we formulate a rule of procedure.

2. Rule of procedure. Test s_1^2/s_2^2 by the F -test. If F is non-significant at some assigned significance level, we use $(n_1s_1^2 + n_2s_2^2)/(n_1 + n_2)$ as the estimate

of σ_1^2 . If F is significant at some assigned significance level we use s_1^2 as the estimate of σ_1^2 . The estimate of σ_1^2 obtained by this rule of procedure will be denoted by e^* .

3. Object of this investigation. If we follow such a rule of procedure, what will be the bias in our estimate e^* of σ_1^2 ?

4. Derivation of the expected value of e^* . First we wish to find

$$E\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}\right), \quad \text{if } \frac{s_1^2}{s_2^2} < \lambda,$$

where λ is the value on the F -distribution corresponding to some assigned significance level for n_1 and n_2 degrees of freedom.

Let $v_1 = s_1^2$, $v_2 = s_2^2$. Since s_1^2 and s_2^2 are independently distributed, the joint distribution of v_1 and v_2 is

$$c_1 v_1^{\frac{1}{2}n_1-1} v_2^{\frac{1}{2}n_2-1} \exp\left[-\frac{1}{2}\left(\frac{n_1 v_1}{\sigma_1^2} + \frac{n_2 v_2}{\sigma_2^2}\right)\right] dv_1 dv_2,$$

where c_1 is a constant and n_1 and n_2 are the respective degrees of freedom.

Let us make the transformation of variables

$$u_1 = n_1 v_1 + n_2 v_2, \quad 0 < u_1 < \infty$$

$$u_2 = \frac{v_1}{v_2} \quad 0 < u_2 < \lambda,$$

then the expected value, E_1 , of $\frac{u_1}{n_1 + n_2}$ for $u_2 < \lambda$ is given by

$$(n_1 + n_2)E_1 = \frac{c_1}{P(u_2 < \lambda)} \int_0^\lambda \int_0^\infty \frac{u_1^{\frac{1}{2}n_1-1}}{(n_1 u_2 + n_2)^{\frac{1}{2}(n_1+n_2)}} \cdot u_1^{\frac{1}{2}(n_1+n_2)} \exp\left[-\frac{1}{2}\left(\frac{u_1}{n_1 u_2 + n_2} + n_2\left(\frac{n_1 u_2}{\sigma_1^2} + \frac{n_2}{\sigma_2^2}\right)\right)\right] du_1 du_2$$

where $P(u_2 < \lambda)$ is the probability of u_2 being less than λ .

Integrating out u_1 and expressing the result in terms of the incomplete beta function we obtain

$$(4) \quad (n_1 + n_2)E_1 = \frac{n_1 I_{x_0}(\frac{1}{2}n_1 + 1, \frac{1}{2}n_2)\sigma_1^2 + n_2 I_{x_0}(\frac{1}{2}n_1, \frac{1}{2}n_2 + 1)\sigma_2^2}{P(u_2 < \lambda)}$$

where $x_0 = (n_1 \varphi \lambda) / (n_2 + n_1 \varphi \lambda)$, $\varphi = \sigma_2^2 / \sigma_1^2$.

We wish now to find the expected value of s_1^2 when $s_1^2/s_2^2 \geq \lambda$. Again we start with the joint distribution of v_1 and v_2 , given above and this time let

$$\frac{v_2}{v_1} = Y, \quad v_1 = v_1,$$

then the expected value, E_2 , of v_1 when $Y \leq \frac{1}{\lambda}$ is

$$E_2 = \frac{c_1}{P\left(Y \leq \frac{1}{\lambda}\right)} \int_0^{1/\lambda} \int_0^\infty Y^{\frac{1}{2}n_2-1} v_1^{\frac{1}{2}(n_1+n_2)} \exp\left[-\frac{1}{2}v_1\left(\frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2}Y\right)\right] dv_1 dY.$$

Integrating out v_1 as a gamma function, and expressing the results as incomplete beta functions we obtain

$$(5) \quad E_2 = \frac{[1 - I_{x_0}(\frac{1}{2}n_1 + 1, \frac{1}{2}n_2)]\sigma_1^2}{P(Y \leq 1/\lambda)},$$

where

$$x_0 = n_1\varphi\lambda/(n_2 + n_1\varphi\lambda) \text{ as before.}$$

5. Final Results. The probability that we use $(n_1s_1^2 + n_2s_2^2)/(n_1 + n_2)$ is $P(u_2 < \lambda)$. From equation (4) the contribution from this case to the mean value of e^* is

$$\frac{n_1 I_{x_0}(\frac{1}{2}n_1 + 1, \frac{1}{2}n_2)\sigma_1^2 + n_2 I_{x_0}(\frac{1}{2}n_1, \frac{1}{2}n_2 + 1)\sigma_2^2}{n_1 + n_2}.$$

The probability that we use s_1^2 is $P(Y \leq 1/\lambda)$. From equation (5) the contribution from this case is

$$[1 - I_{x_0}(\frac{1}{2}n_1 + 1, \frac{1}{2}n_2)]\sigma_1^2.$$

The expected value of e^* is obtained by combining the two cases, i.e.,

$$(6) \quad E(e^*) = \left[1 + \frac{n_2}{n_1 + n_2} \left\{ I_{x_0}(\frac{1}{2}n_1, \frac{1}{2}n_2 + 1) \frac{\sigma_2^2}{\sigma_1^2} - I_{x_0}(\frac{1}{2}n_1 + 1, \frac{1}{2}n_2) \right\}\right] \sigma_1^2.$$

Hence the bias in e^* , expressed as a fraction of σ_1^2 is

$$(7) \quad \frac{n_1}{n_1 + n_2} \left[I_{x_0}(\frac{1}{2}n_1, \frac{1}{2}n_2 + 1) \frac{\sigma_2^2}{\sigma_1^2} - I_{x_0}(\frac{1}{2}n_1 + 1, \frac{1}{2}n_2) \right]$$

We note that in estimating σ_1^2 there will be a positive bias, no bias, or a negative bias according as

$$\frac{I_{x_0}(\frac{1}{2}n_1, \frac{1}{2}n_2 + 1)}{I_{x_0}(\frac{1}{2}n_1 + 1, \frac{1}{2}n_2)}$$

is greater than, equal to, or less than σ_1^2/σ_2^2 .

6. Identity and checks. If $\sigma_1^2 = \sigma_2^2$, then in section 4, $E_1 = \sigma_1^2$ and

$$P(u_2 < \lambda) = I_{x_0}(\frac{1}{2}n_1, \frac{1}{2}n_2).$$

From (4) this gives the identity

$$(n_1 + n_2)I_{x_0}(\frac{1}{2}n_1, \frac{1}{2}n_2) = n_1I_{x_0}(\frac{1}{2}n_1 + 1, \frac{1}{2}n_2) + n_2I_{x_0}(\frac{1}{2}n_1, \frac{1}{2}n_2 + 1)$$

where $x_0 = n_1\lambda/(n_2 + n_1\lambda)$. This identity may be established easily by elementary calculus.

The first result in equation (6) may be checked by noting that when $\lambda = \infty$, i.e. when the two mean squares are always pooled, x_0 is 1 and equation (6) reduces to $(n_1\sigma_1^2 + n_2\sigma_2^2)/(n_1 + n_2)$. Similarly when $\lambda = 0$, in which case there is no pooling, $x_0 = 0$, and equation (6) reduces to σ_1^2 .

7. Discussion. In making a choice of an appropriate estimate of σ_1^2 we may consider three procedures:

(1) Use s_1^2 always. This has the merit of having no bias, but is likely to have a large sampling error.

(2) Always pool, i.e., use $\frac{n_1s_1^2 + n_2s_2^2}{n_1 + n_2}$. When $\sigma_1^2 \neq \sigma_2^2$ this is biased, but in compensation will have less sampling error than (1) since it will be based on $(n_1 + n_2)$ degrees of freedom.

(3) Use the test of significance of $\frac{s_1^2}{s_2^2}$ as a criterion in making the decision as to whether to pool the two mean squares or not. If the test discriminates properly between cases where pooling should or should not be made, the preliminary test of significance criterion will utilize the extra n_2 degrees of freedom whenever permissible and also avoid the bias in method (2).

In Table I the expected value $E(c^*)$ divided by σ_1^2 , is given for two sets of values of n_1, n_2 somewhat typical of those frequently encountered in applied work, and for a series of values of σ_2^2/σ_1^2 . In addition to the case of always pooling ($\lambda = \infty$) and that of never pooling ($\lambda = 0$), the results for λ at the 1 percent, 5 percent, 20 percent levels and for $\lambda = 1$ have been tabulated. By subtracting unity from the results the bias is obtained as a fraction of σ_1^2 . The Table was computed from the incomplete beta function Tables [6].

When the two mean squares are always pooled, the fractional bias is negative and increases numerically as σ_2^2 becomes small relative to σ_1^2 . By examination of the values in Table I for $\sigma_2^2/\sigma_1^2 = .1$, it will be seen that the preliminary test of significance controls the bias well when σ_2^2 is much smaller than σ_1^2 , that is when a large bias from pooling is most to be feared. This result happens because in such cases the preliminary test allows pooling only in a small proportion of samples.

If λ is taken at the 1 or 5 percent levels, the maximum bias appears to occur when σ_2^2/σ_1^2 is in the region 0.4-0.5, there being little bias when σ_2^2 is near σ_1^2 . The lower values of λ (20 percent or λ equals 1) control the bias satisfactorily in the region $\sigma_2^2 < .6\sigma_1^2$, but have a fairly substantial positive bias when $\sigma_2^2 = \sigma_1^2$, that is when pooling would actually be justified. By use of the relation between the incomplete beta function and the sum of the terms of a binomial series it

can be shown that there is always a positive bias when $\sigma_1^2 = \sigma_2^2$ and that for given numbers of degrees of freedom this bias is greatest when $\lambda = 1$.

To summarize from the example in Table I, it seems that for small values of n_1 and n_2 none of the values of λ which have been investigated controls the bias throughout the whole range $0 \leq \sigma_2^2/\sigma_1^2 \leq 1$.

TABLE I
Expected Value of c^/σ_1^2 : $E(c^*)/\sigma_1^2$*

Case 1. $n_1 = 4, n_2 = 20$										
	σ_2^2/σ_1^2									
	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\lambda = \infty$.250	.333	.417	.500	.583	.667	.750	.833	.917	1.00
$\lambda_{.01} = 4.43$.965	.870	.791	.750	.748	.775	.821	.880	.948	1.02
$\lambda_{.05} = 2.87$.991	.960	.924	.901	.892	.903	.930	.970	1.02	1.08
$\lambda_{.20} = 0.00$	1.00	.999	1.00	1.01	1.02	1.04	1.07	1.11	1.15	1.20
$\lambda = 1$	1.00	1.00	1.01	1.03	1.05	1.08	1.11	1.15	1.20	1.25
$\lambda = 0$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Case 2: $n_1 = 12, n_2 = 10$										
	σ_2^2/σ_1^2									
	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\lambda = \infty$.591	.636	.682	.727	.773	.818	.864	.909	.955	1.00
$\lambda_{.01} = 4.71$.981	.896	.833	.814	.824	.850	.884	.922	.963	1.00
$\lambda_{.05} = 2.91$.998	.973	.935	.909	.901	.908	.928	.955	.989	1.03
$\lambda_{.20} = 0.00$	1.00	.998	.993	.987	.986	.991	1.00	1.02	1.04	1.07
$\lambda = 1$	1.00	1.00	1.00	1.00	1.01	1.02	1.04	1.06	1.08	1.11
$\lambda = 0$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

8. The variance of c^* . Using the same method we may obtain the variance of c^* . The final result is

$$\begin{aligned}
 V = & \frac{n_1(n_1 + 2)I_{x_0}(\tfrac{1}{2}n_1 + 2, \tfrac{1}{2}n_2)\sigma_1^4 + 2n_1n_2I_{x_0}(\tfrac{1}{2}n_1 + 1, \tfrac{1}{2}n_2 + 1)\sigma_1^2\sigma_2^2}{(n_1 + n_2)^2} \\
 & + \frac{n_2(n_2 + 2)I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2 + 2)\sigma_2^4}{(n_1 + n_2)^2} \\
 (8) \quad & + \frac{n_1 + 2}{n_1} [1 - I_{x_0}(\tfrac{1}{2}n_1 + 2, \tfrac{1}{2}n_2)]\sigma_1^4 - \left[1 + \frac{n_2}{n_1 + n_2} \right. \\
 & \quad \left. \left\{ I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2 + 1) \frac{\sigma_2^2}{\sigma_1^2} - I_{x_0}(\tfrac{1}{2}n_1 + 1, \tfrac{1}{2}n_2) \right\} \right] \sigma_1^4.
 \end{aligned}$$

From the relations in deriving this result the following identity was obtained:

$$(n_1 + n_2 + 2)(n_1 + n_2)I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2) = n_1(n_1 + 2)I_{x_0}(\tfrac{1}{2}n_1 + 2, \tfrac{1}{2}n_2) \\ + 2n_1n_2I_{x_0}(\tfrac{1}{2}n_1 + 1, \tfrac{1}{2}n_2 + 1) + n_2(n_2 + 2)I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2 + 2).$$

This identity can be readily established by elementary calculus.

As a check on the result in equation (8), we note that if $\lambda = \infty$, then $x_0 = 1$, and $\frac{s_1^2}{s_2^2} < \lambda$ always. The variance of the estimate of variance becomes

$2(n_1s_1^4 + n_2s_2^4)/(n_1 + n_2)^2$, which checks with the variance of $(n_1s_1^2 + n_2s_2^2)/(n_1 + n_2)$ for the case of always pooling. If in addition $\sigma_1^2 = \sigma_2^2$, then $V = 2\sigma_1^4/(n_1 + n_2)$. If $\lambda = 0$, then $x_0 = 0$, and $s_1^2/s_2^2 \geq \lambda$ always. The variance of the estimate of variance becomes $2\sigma_1^4/n_1$ which checks with the variance of s_1^2 for the case of never pooling.

The expression for the variance of e^* enables us to investigate how much has been gained in terms of reduction in variance by the use of the preliminary test. The quantity $\{V + (\text{Bias})^2\}$ is the appropriate value for the whole sampling error, where V is given by (8) and the bias by (7). For the two numerical examples these quantities are shown as fractions of σ_1^4 in Table II.

As a standard of comparison the variances of the estimate s_1^2 (no pooling) will be used. In these examples the preliminary test with $\lambda = 1$ produces a variance smaller than that of s_1^2 for all values of σ_2^2/σ_1^2 except the lowest (0.1) where the two variances are equal. As λ is taken successively higher there is a substantial reduction in variance when σ_2^2 is near σ_1^2 but an increase in variance over that of s_1^2 when σ_2^2/σ_1^2 is small. Throughout nearly all the range of values of σ_2^2/σ_1^2 , the smallest variance is obtained by always pooling ($\lambda = \infty$), despite the relatively large bias given by that method. This result is a reflection of the instability of estimates of variance which are based on only a few degrees of freedom.

III. EXAMPLE TWO: TEST OF A REGRESSION COEFFICIENT

1. Regression and some properties of orthogonal functions. Let

$$y = \beta_1x_1 + \beta_2x_2 + e$$

be a linear regression of y on the two variates x_1 and x_2 in which β_1 and β_2 are the respective population regression coefficients and e is the error. We assume that x_1 , x_2 and y are measured from their respective sample means and that the values of x_1 and x_2 are fixed from sample to sample. In order to make comparisons among samples of different sizes we assume that x_1 and x_2 have unit variances and correlation coefficient¹ ρ so that

$$S(x_1^2) = n - 1, \quad S(x_2^2) = n - 1, \quad S(x_1x_2) = \rho(n - 1),$$

¹ Although ρ is commonly used to denote a population correlation coefficient, we are using it here for the sample correlation coefficient between the fixed variates x_1 and x_2 .

TABLE II

The Variance of c^* About its True Mean $\cdot \frac{V + (\text{Bias})^2}{\sigma_1^4}$

Case 1. $n_1 = 4, n_2 = 20$										
	σ_2^2/σ_1^2									
	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\lambda = \infty$.577	.462	.360	.275	.205	.149	.111	.087	.076	.084
$\lambda_{.01} = 4.43$.560	.620	.603	.523	.350	.323	.243	.184	.150	.137
$\lambda_{.05} = 2.87$.514	.545	.554	.528	.479	.414	.353	.299	.260	.237
$\lambda_{.20} = 0.00$.501	.500	.493	.480	.458	.435	.408	.374	.367	.360
$\lambda = 1$.499	.493	.480	.462	.441	.423	.401	.389	.381	.387
$\lambda = 0$.500	.500	.500	.500	.500	.500	.500	.500	.500	.500

Case 2: $n_1 = 12, n_2 = 10$										
	σ_2^2/σ_1^2									
	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\lambda = \infty$.217	.183	.154	.130	.112	.097	.088	.084	.085	.091
$\lambda_{.01} = 4.71$.185	.218	.203	.171	.141	.118	.103	.094	.092	.096
$\lambda_{.05} = 2.91$.170	.187	.194	.183	.163	.142	.125	.114	.109	.109
$\lambda_{.20} = 0.00$.167	.169	.171	.170	.164	.156	.146	.139	.135	.135
$\lambda = 1$.167	.166	.165	.162	.158	.152	.148	.144	.144	.147
$\lambda = 0$.167	.167	.167	.167	.167	.167	.167	.167	.167	.167

where $S(x_1^2)$ denotes summation of x_1^2 over the sample, with similar meanings for $S(x_2^2)$ and $S(x_1x_2)$, where n is the sample size.

We make the orthogonal transformations

$$\xi_1 = x_1, \quad \xi_2 = x_2 - \rho x_1,$$

then

$$y = \beta_1 \xi_1 + \beta_2 (\xi_2 + \rho \xi_1) + e.$$

But

$$S\xi_1^2 = n - 1, \quad S\xi_2^2 = (n - 1)(1 - \rho^2), \quad S(\xi_1\xi_2) = 0,$$

therefore

$$S(y\xi_1) = \beta_1(n - 1) + \beta_2\rho(n - 1) + S(x_1e),$$

and

$$S(y\xi_2) = \beta_2(n - 1)(1 - \rho^2) + S(x_2 - \rho x_1)e.$$

Now if we represent the regression coefficients of y on the ξ 's as B 's we have

$$B_1 S(\xi_1^2) = S(\xi_1 y), \quad B_2 S(\xi_2^2) = S(\xi_2 y).$$

The reduction in the total sum of squares due to x_1 ignoring x_2 is

$$B_1 S(y\xi_1) = \frac{[S(\xi_1 y)]^2}{S(\xi_1^2)} = \frac{[(n-1)(\beta_1 + \beta_2 \rho) + S(x_1 e)]^2}{n-1}.$$

The reduction in the total sum of squares due to x_2 after fitting x_1 is

$$B_2 S(y\xi_2) = \frac{[S(\xi_2 y)]^2}{S(\xi_2^2)} = \frac{[(n-1)\beta_2(1-\rho^2) + S(x_2 - \rho x_1)e]^2}{(n-1)(1-\rho^2)}.$$

The reduction in the total sum of squares due to regression is

$$\frac{[(n-1)(\beta_1 + \beta_2 \rho) + S(x_1 e)]^2}{n-1} + \frac{[S(x_2 - \rho x_1) + (n-1)\beta_2(1-\rho^2)]^2}{(n-1)(1-\rho^2)},$$

in which the two parts are independently distributed.

Let b'_1 be the regression coefficient of y on x_1 when the term containing x_2 is omitted from the regression equation. Now,

$$b'_1 = B_1 = \frac{S(\xi_1 y)}{S(\xi_1^2)} = \frac{(n-1)(\beta_1 + \beta_2 \rho) + S(x_1 e)}{n-1}.$$

Hence

$$(9) \quad E(b'_1) = \beta_1 + \beta_2 \rho.$$

Let b_2 be the regression coefficient of y on x_2 if both x_1 and x_2 are used. Then

$$b_2 = B_2 = \frac{S(\xi_2 y)}{S(\xi_2^2)} = \frac{(n-1)\beta_2(1-\rho^2) + S(x_2 - \rho x_1)e}{(n-1)(1-\rho^2)}.$$

And

$$V(b_2) = \frac{S(\xi_2^2)}{[S(\xi_2^2)]^2} = \frac{1}{S(\xi_2^2)} = \frac{1}{(n-1)(1-\rho^2)}.$$

The normal equations for $Y = b_1 x_1 + b_2 x_2$ are

$$b_1 S(x_1^2) + b_2 S(x_1 x_2) = S(x_1 y),$$

$$b_1 S(x_1 x_2) + b_2 S(x_2^2) = S(x_2 y).$$

Now

$$b'_1 = \frac{S(x_1 y)}{S(x_1^2)} = b_1 + b_2 \frac{S(x_1 x_2)}{S(x_1^2)}.$$

Therefore

$$b'_1 = b_1 + b_2 \rho,$$

or

$$b_1 = b'_1 - b_2 \rho.$$

Therefore

$$(10) \quad E(b_1) = \beta_1 + \beta_2\rho - \rho E(b_2).$$

We notice that if $\rho = 0$, b_1 is unbiased in any selected portion of the population

2. Statement of the problem. To obtain an estimate of b_1 , in a particular analysis in hand, in which it is desirable to choose by means of a test of significance between using the regression equation $Y = b_1x_1 + b_2x_2$ and $Y' = b'_1x_1$, we formulate a rule of procedure.

3. Rule of procedure. Calculate the following analysis of variance:

	Degrees of freedom	Sum of squares	Mean square
Reduction due to x_1	1	$\frac{[(n-1)(\beta_1 + \beta_2\rho) + S(x_1e)]^2}{n-1}$	s_1^2
Reduction due to x_2 after fitting x_1	1	$\frac{[(n-1)\beta_2(1-\rho^2) + S(x_2 - \rho x_1)e]^2}{(n-1)(1-\rho^2)}$	s_2^2
Residual	$n-3$	$S(y - Y)^2$	s_3^2

Test $\frac{s_2^2}{s_3^2}$ by the F -test. If F is non-significant at some assigned significance level we omit the term containing x_2 and use

$$b'_1 = \frac{(n-1)(\beta_1 + \beta_2\rho) + S(x_1e)}{n-1}$$

as the estimate of β_1 . If F is significant, we retain the term containing x_2 , and use b_1 as the estimate of β_1 . The estimate obtained by this rule will be called b^* .

4. Object of this investigation. If we follow such a rule of procedure, what will be the bias in b^* as an estimate of β_1 ?

5. Mathematical derivation of the bias. First, we wish $E(b'_1)$ when

$$\frac{s_2^2}{s_3^2} < \lambda \quad \text{or} \quad \frac{b_2^2}{s_3^2} < \frac{\lambda}{(1-\rho^2)(n-1)},$$

where λ is the value on Snedecor's F -distribution corresponding to some assigned significance level for 1 and $(n-3)$ degrees of freedom. From (9) we have

$$(11) \quad E(b'_1) = \beta_1 + \beta_2\rho,$$

no matter what the value of $\frac{s_2^2}{s_3^2}$, since from section 1, s_1^2 and s_2^2 are independently distributed.

Next we wish $E(b_1)$ when $\frac{s_2^2}{s_3^2} \leq \lambda$ or $\frac{b_2^2}{s_3^2} \leq \frac{\lambda}{(1-\rho^2)(n-1)}$. To obtain this we find it more convenient to find first $E(b_2)$ when

$$\frac{b_2^2}{s_3^2} \leq \frac{\lambda}{(1-\rho^2)(n-1)}, \quad \text{or} \quad \frac{v_3}{b_2^2} \leq \frac{1}{\lambda c_{22}},$$

where

$$v_3 = s_3^2 \quad \text{and} \quad c_{22} = \frac{1}{(n-1)(1-\rho^2)}.$$

The joint distribution of b_2, v_3 is

$$K e^{-(b_2 - \beta_2)^2 / 2c_{22}} v_3^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}v_3} dv_3 db_2,$$

where K is a constant. We make the transformation of variables

$$u = \frac{v_3}{b_2^2}, \quad dv_3 = b_2^2 du;$$

then the joint distribution of b_2 and u is

$$K e^{-(b_2 - \beta_2)^2 / 2c_{22}} (u b_2^2)^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}u} b_2^2 du db_2.$$

Taking the expected value when $u \leq \frac{1}{\lambda c_{22}}$ we have

$$E(b_2) = \frac{K}{P\left(u \leq \frac{1}{\lambda c_{22}}\right)} \cdot \int_{-\infty}^{\infty} \int_0^{1/\lambda c_{22}} b_2 |b_2|^{n-3} u^{\frac{1}{2}(n-3)} \exp\left[-\frac{(b_2 - \beta_2)^2}{2c_{22}} - \frac{(n-3)}{2} b_2^2 u\right] du db_2,$$

where $v_2 = s_2^2$, and $P\left(u \leq \frac{1}{\lambda c_{22}}\right)$ is the probability that u be less than or equal to $\frac{1}{\lambda c_{22}}$.

Dropping subscripts for convenience and expanding the factor which involves e to the first power of b , we have

$$E(b) = \frac{K e^{-\beta^2/2c}}{P\left(u \leq \frac{1}{\lambda c}\right)} \int_{-\infty}^{\infty} \int_0^{1/\lambda c} b |b|^{n-3} u^{\frac{1}{2}(n-3)} \cdot \exp\left\{-\frac{1}{2}b^2\left[1 + \frac{(n-3)cu}{c}\right]\right\} \left[1 + \left(\frac{b\beta}{c}\right) + \frac{1}{2!}\left(\frac{b\beta}{c}\right)^2 + \dots\right] du db,$$

where

$$-\infty < b < \infty, \quad 0 < u < \frac{1}{\lambda c}.$$

Now, clearly the even terms of the series vanish whether n is odd or even when b is integrated out.

After integration with respect to b , we have an infinite series in which the typical term (apart from constants) is of the form

$$u^{1(n-b)}/[1 + (n-3)cu]^{1(n+r)}$$

where r is an even positive integer. Subsequent integration with respect to u leads to an infinite series of incomplete integrals of the F distribution. By transforming the integrals, the series may be expressed in terms of incomplete beta functions as follows:

$$E(b_2) = \frac{\beta_2 e^{-\beta_2^2/2c_{22}}}{P\left(u \leq \frac{1}{\lambda c_{22}}\right)} \cdot \left[I_{x_0}\left(\frac{n-3}{2}, \frac{3}{2}\right) + \left(\frac{\beta_2^2}{2c_{22}}\right) I_{x_0}\left(\frac{n-3}{2}, \frac{5}{2}\right) + \frac{1}{2!} \left(\frac{\beta_2^2}{2c_{22}}\right)^2 I_{x_0}\left(\frac{n-3}{2}, \frac{7}{2}\right) + \dots \right].$$

Let

$$a = \frac{\beta_2^2}{2c_{22}} \quad \text{or} \quad a = \frac{(1-\rho^2)(n-1)\beta_2^2}{2}.$$

Then we have

$$(12) \quad E(b_2) = \frac{\beta_2}{P\left(u \leq \frac{1}{\lambda c_{22}}\right)} \sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0}\left(\frac{n-3}{2}, \frac{3}{2} + i\right),$$

$$\text{where } x_0 = \frac{1}{\frac{\lambda}{n-3} + 1},$$

and λ is the desired % point of the F -distribution for 1 and $(n-3)$ degrees of freedom. Now from (10) we have

$$E(b_1) = \beta_1 + \beta_2 \rho - \rho E(b_2)$$

which enables us to obtain $E(b_1)$ from (12).

6. Final result. From (10), (11) and (12) we have

$$\begin{aligned} E(b^*) &= P\left(\frac{v_2}{v_3} < \lambda\right) (\beta_1 + \beta_2 \rho) + \left[1 - P\left(\frac{v_2}{v_3} < \lambda\right)\right] [\beta_1 + \rho\{\beta_2 - E(b_2)\}] \\ &= \beta_1 + \rho\beta_2 - \left[1 - P\left(\frac{v_2}{v_3} < \lambda\right)\right] \rho E(b_2). \end{aligned}$$

The bias in b^* is

$$\rho \left[\beta_2 - \left\{ 1 - P \left(\frac{v_2}{v_3} < \lambda \right) \right\} E(b_2) \right].$$

Substituting the value of $E(b_2)$, we obtain

$$\text{Bias} = \rho \beta_2 \left[1 - \sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0} \left(\frac{n-3}{2}, \frac{3}{2} + i \right) \right],$$

where $x_0 = \frac{1}{\frac{\lambda}{n-3} + 1}$, $a = (1 - \rho^2) \binom{n-1}{2} \beta_2^2$.

7. Checks. From section 5 we have

$$E(b_2) = \frac{\beta_2}{P \left(\frac{v_2}{v_3} \geq \lambda \right)} \sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0} \left(\frac{n-3}{2}, \frac{3}{2} + i \right),$$

where $x_0 = \frac{1}{\frac{\lambda}{n-3} + 1}$.

If $\lambda = 0$, then $x_0 = 1$, and $E(b_2) = \beta_2$.

Also from section 5 we have

$$\text{Bias} = \rho \beta_2 \left[1 - \sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0} \left(\frac{n-3}{2}, \frac{3}{2} + i \right) \right].$$

If $\lambda = 0$, then $x_0 = 1$, and $\text{Bias} = 0$.

If $\lambda = \infty$, then $x_0 = 0$, and $\text{Bias} = \rho \beta_2$.

8. Discussion. From the mathematical form of the bias,

$$\text{Bias} = \rho \beta_2 \left[1 - \sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0} \left(\frac{n-3}{2}, \frac{3}{2} + i \right) \right],$$

where $x_0 = \frac{1}{\frac{\lambda}{n-3} + 1}$,

four deductions follow immediately: (i) There is no bias in estimating β_1 , if ρ or β_2 is zero. (ii) The coefficient of β_2 in the formula is less than or at most equal to one in absolute value. (iii) The sign of the bias depends upon the signs of ρ and β_2 ; it is positive if both are positive or both negative, it is negative if ρ and β_2 have opposite signs. (iv) The bias in estimating β_1 is independent of β_1 .

We shall discuss the bias in a few special cases by means of selected values of n , ρ , β_2 and λ . In Table III are exhibited the values of the bias for n equal to 5, 11, 21, each at ρ equal to .2, .4, .6, .8, and β_2 equal to .1, .4, 1.0, 2.0,

and 4.0 These values have been computed at the 5% point for λ , and at $\lambda = 1$. These special cases seem to indicate: (i) If we fix ρ , β_2 , and λ and increase n , then the bias decreases. (ii) If we fix ρ , β_2 , and n and change λ from the 5% point to $\lambda = 1$, the bias decreases considerably. (iii) If we fix ρ , n , λ and increase

TABLE III
The Bias in Estimating β_1

$\rho \backslash \beta_2$	$\lambda_{.05} = 18.513$ $n = 5$				$\lambda_{.05} = 5.318$ $n = 11$				$\lambda_{.05} = 4.414$ $n = 21$			
	.2	.4	.6	.8	.2	.4	.6	.8	.2	.4	.6	.8
0.1	.017	.034	.051	.069	.015	.030	.046	.061	.014	.029	.044	.059
0.4	.067	.134	.202	.272	.049	.101	.159	.227	.033	.071	.122	.193
1.0	.142	.292	.455	.640	.028	.072	.164	.350	.001	.006	.025	.132
2.0	.162	.358	.627	1.038	.000	.000	.001	.083	.000	.000	.000	.001
4.0	.035	.101	.282	.898	.000	.000	.000	.000	.000	.000	.000	.000
$\rho \backslash \beta_2$	$\lambda = 1$ $n = 5$				$\lambda = 1$ $n = 11$				$\lambda = 1$ $n = 21$			
	.2	.4	.6	.8	.2	.4	.6	.8	.2	.4	.6	.8
0.1	.004	.008	.011	.015	.004	.008	.011	.015	.004	.008	.011	.015
0.4	.012	.026	.040	.057	.009	.019	.032	.051	.005	.011	.022	.041
1.0	.011	.025	.049	.095	.001	.003	.010	.039	.000	.000	.001	.009
2.0	.000	.002	.008	.043	.000	.000	.000	.000	.000	.000	.000	.000
4.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

β_2 the bias increases and then decreases. This may be explained in the following manner. From section 6, the bias may be written in the form

$$\text{Bias} = \rho\beta_2 \frac{P\left(\frac{v_2}{v_3} < \lambda\right)}{P\left(\frac{v_2}{v_3} > \lambda\right)} \sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0}\left(\frac{n-3}{2}, \frac{3}{2} + i\right).$$

Now if ρ , n , λ are held constant and β_2 is relatively small, $P\left(\frac{v_2}{v_3} < \lambda\right)$ is relatively large and $\sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0}\left(\frac{n-3}{2}, \frac{3}{2} + i\right)$ is relatively large, but $P\left(\frac{v_2}{v_3} > \lambda\right)$ is relatively small. Hence, for a while as we increase β_2 the bias will increase, but as β_2 gets larger $P\left(\frac{v_2}{v_3} < \lambda\right)$ and $\sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0}\left(\frac{n-3}{2}, \frac{3}{2} + i\right)$ becomes smaller while $P\left(\frac{v_2}{v_3} > \lambda\right)$ becomes larger. Hence, a value of β_2 will be reached at which the

bias begins to decrease. (iv) If we fix n , d_2 , and λ and increase ρ , the bias increases without exception.

The above results were obtained under the assumption that a test of significance criterion is used in making a choice as to the number of independent variables to be retained after the regression $y = b_1x_1 + b_2x_2$ has been fitted. If this test of significance criterion is used, we may wish to have a means of controlling the bias. From a study of Table III we note that the bias may be decreased by increasing n and by using $\lambda = 1$. We also notice that as β_2 increases from 0.1 to 4.0 the bias increases and then decreases; and so passes through a maximum value. Hence, if we have a regression in which β_2 is fairly well below or above this maximum value, we would expect a smaller bias.

The bias in estimating β_1 is "unstudentized," i.e., is a function of the population parameters ρ and β_2 . In any particular analysis in hand, it would be necessary to know the values of ρ and β_2 or be willing to use estimates of them obtained from the data.

It is realized that only a beginning has been made on the regression problem: an investigation should be undertaken of the more general problem of the use of a test of significance criterion in making a choice as to the number of independent variables to be retained after the regression

$$y = b_1x_1 + b_2x_2 + \cdots + b_nx_n$$

has been fitted.

REFERENCES

- [1] J. WISHART and A. R. CLAPHAM, "A study in sampling techniques: the effect of artificial fertilizers on the yield of potatoes," *Jour. Agri. Sci.*, Vol. 19 (1929), Part 4, p. 605.
- [2] W. G. COCHRAN, "The omission or addition of an independent variate in multiple linear regression," *Jour. Roy. Stat. Soc. Suppl.*, Vol. 5 (1938), p. 171.
- [3] G. W. SNEDECOR, *Statistical Methods*, Third edition. Iowa State College Press, Ames, Iowa. 1940 Sec. 14.3.
- [4] M. S. BARTLETT, "Square root transformation in analysis of variance," *Jour. Roy. Stat. Soc. Suppl.*, Vol. 3 (1936), pp. 76-77.
- [5] JOSEPH BERKSON, "Tests of significance considered as evidence," *Jour. Amer. Stat. Assn.*, Vol. 37 (1942), pp. 325-335.
- [6] KARL PEARSON (Editor), *Tables of the Incomplete Beta Function*, Cambridge Univ. Press. *Biometrika* Univ. Coll., London, 1934.

THE PROBABILITY OF CONVERGENCE OF AN ITERATIVE PROCESS OF INVERTING A MATRIX

BY JOSEPH ULLMAN

Columbia University

Introduction. The inversion of a matrix is a computational problem of wide application. This is a further study of an efficient iterative method of matrix inversion described by Harold Hotelling [1], with an examination of the probability of convergence in relation to the accuracy of the initial approximation. The lines of investigation were suggested both by his article and by helpful comments made during the course of the research.

The inverse of a matrix can be obtained to any desired degree of accuracy by using a variation of the Doolittle method, and starting with a sufficient number of accurate decimal places in the matrix being inverted. This procedure becomes inefficient if the order of the matrix is large, or if the desired degree of accuracy is very great. In either case the efficiency can be greatly increased by first obtaining an approximation to a small number of decimal places and then applying a method of iteration until the desired accuracy is achieved.

1. Iterative methods. Hotelling's method of iteration is as follows. Let A be the matrix to be inverted and let C_0 be the approximation to the inverse. Calculate in turn C_1, C_2, \dots where,

$$(1.1) \quad C_{m+1} = C_m(2 - AC_m).$$

This sequence of matrices will converge to the inverse of A if the roots of

$$(1.2) \quad D = 1 - AC_0,$$

are all less than unity in absolute value.

The iterative method (1.1) will be generalized to yield a class of iterative methods, one element of which will be shown to be more efficient, in certain cases, than method (1.1). The generalized iterative method is,

$$(1.3) \quad C_{m+1} = C_m\{1 + (1 - AC_m) + (1 - AC_m)^2 + \dots + (1 - AC_m)^{k-1}\}.$$

For every k , the condition for convergence is that the roots of the matrix (1.2) all be less than unity in absolute value.

A method of comparing the efficiency of these different iterative methods arises from the following considerations. Since

$$(1.4) \quad C_0 = A^{-1}(AC_0),$$

which is equivalent to

$$(1.5) \quad C_0 = A^{-1}(1 - D),$$

it follows that

$$(1.6) \quad A^{-1} = C_0(1 - D)^{-1}.$$

When the roots of D are all less than unity in absolute value, (1.6) has the infinite expansion,

$$(1.7) \quad A^{-1} = C_0(1 + D + D^2 + D^3 + \cdots).$$

The general iterative process (1.3) generates the infinite series in the following manner,

$$(1.8) \quad (1 + D + D^2 + \cdots + D^{k-1})(1 + D^k + \cdots + D^{k^2-k}) \cdots \\ \cdot (1 + D^{k^2} + \cdots + D^{k^2(k-1)}) \cdots$$

Each parentheses corresponds to one iteration. Hence k^m terms are generated by m iterations. In order to achieve the accuracy of n terms in (1.7), $m = \log_e n / \log_e k$ iterations are required. Each iteration involves k matrix multiplications, so that $km = k \log_e n / \log_e k$ is the total number of matrix multiplications necessary to achieve this degree of accuracy.

The integer for which this is a minimum is three. Therefore the "most efficient" method of iteration is,

$$(1.10) \quad C'_{m+1} = C'_m \{1 + (1 - AC'_m) + (1 - AC'_m)^2\}.$$

If the desired degree of accuracy can be achieved by one application of (1.1), or by two applications of (1.1) but not by one application of (1.10), then (1.1) is preferable.

2. The condition for convergence. The sequence,

$$(2.1) \quad C_1, C_2, C_3, \cdots$$

obtained from (1.3) will converge to the inverse of A if the roots of

$$(2.2) \quad D = 1 - AC_0,$$

are all less than unity in absolute value. The following assumptions determine the nature of D .

We assume that the expected value of each element of the first approximation C_0 is equal to the corresponding element of the exact inverse of A . The actual values of the elements of C_0 will deviate from their expected values. We will consider two important cases. If the deviations are entirely due to the fact that the elements of C_0 are only accurate to a limited number of decimal places, say k , then the deviations may be regarded as distributed with constant density over a range of length 10^{-k} . It will be assumed that the deviations of the elements of C_0 from their expected values are independent. While this case arises in practice, we will first treat a closely related case, which lends itself to exact treatment more readily. We assume that the deviations of the elements of C_0

are normally distributed about their expected values, with variance $\mu = 10^{-2k}/12$. The variance μ is the same as that which arises if the probability density is uniform with range 10^{-k} .

The elements of E , the matrix of deviations,

$$(2.3) \quad E = A^{-1} - C_0, .$$

are independently and normally distributed. Combining (2.2) and (2.3) we obtain

$$(2.4) \quad D = 1 - AC_0 = A(A^{-1} - C_0) = AE.$$

Let p be the order of the matrix A . Each element of D will be a linear combination of p independently and normally distributed variables, and therefore will itself be normally distributed. A sufficient condition for all the roots of D to be less than unity in absolute value, and hence for the process of iteration to converge, is for the sum of the squares of the elements of D to be less than unity in absolute value. We will use the following notation

(2.5) d_{ij} : the element of D in the i th row and j th column,

$$N_D^2 = \sum_i \sum_j d_{ij}^2$$

A procedure suggested by this relationship is to determine the probability distribution of N_D^2 , so that probability statements concerning the absolute value of the roots of D can be made. Because the elements of D are not all independent, no multiple of N_D^2 can be expected to have the $\chi^2_{(p^2)}$ distribution.¹

The distribution of N_D^2 is shown to be closely related to the chi-square distribution in the next section, and on the basis of this relationship, lower bounds to the probability of convergence of the iterative process are developed in section 4. In section 5 the exact distribution of the norm is obtained for a general class of cases. The final section is concerned with the validity of applying the results of this study to a practical situation, where the deviations of the elements of C_0 from their expected values are uniformly, rather than normally, distributed.

3. An equivalence. Let e_{ij} be the element of E in the i th row and the j th column, and a_{ij} be the element of A in the i th row and the j th column. From (2.4) and (2.5), we find that

$$d_{ij} = \sum_k a_{ik} e_{kj} .$$

Since the elements of E are independently and normally distributed with variance $\mu = 10^{-2k}/12$ it follows readily that

$$(3.2) \quad E[e_{ij}e_{kh}] = \delta_{ik}\delta_{jh}\mu$$

¹ The number in the parentheses will indicate the number of degrees of freedom of the chi-square distribution.

Making use of (3.1) and (3.2), we find that for two $d_{i,j}$ in the same column,

$$(3.3) \quad E[d_{ij}, d_{kj}] = \mu \sum_i a_{it} a_{kt},$$

while for any two $d_{i,j}$ in different columns,

$$(3.4) \quad E[d_{ij}, d_{kh}] = 0.$$

From (3.3) and (3.4) it follows that it is permissible to regard the elements of the p columns of D as the coordinates of p independently selected points from a multivariate normal universe with covariance matrix $\sigma = \mu A A'$. We will let $\lambda = \sigma^{-1}$.

The moment generating function of the sum of squares of the coordinates of any point is

$$(3.5) \quad \frac{|\lambda|^{1/2}}{|\lambda - 2t|^{1/2}}.$$

This can also be written as

$$(3.6) \quad \frac{1}{(1 - 2\sigma_1 t)^{1/2} (1 - 2\sigma_2 t)^{1/2} \cdots (1 - 2\sigma_p t)^{1/2}},$$

where $\sigma_1, \dots, \sigma_p$ are the characteristic roots of σ .

Since N_p^2 is the sum of p independent expressions of this type, its moment generating function is the p th power of (3.6),

$$(3.7) \quad \frac{1}{(1 - 2\sigma_1 t)^{1/2} \cdots (1 - 2\sigma_p t)^{1/2}}.$$

This expression is the moment generating function of

$$(3.8) \quad \sigma_1 \chi_{(p)1}^2 + \sigma_2 \chi_{(p)2}^2 + \cdots + \sigma_p \chi_{(p)p}^2,$$

where the $\chi_{(p)i}^2$ are all independent.

Writing the roots as

$$(3.9) \quad \sigma_0, \sigma_0 - k_1, \dots, \sigma_0 - k_{p-1},$$

where σ_0 is the largest root of σ , and all $k_i > 0$, it follows that N_p^2 has the same distribution as

$$(3.10) \quad \sigma_0 \sum_i \chi_{(p)i}^2 - \sum_{j=1}^{p-1} k_j \chi_{(p)j}^2.$$

Therefore, making use of the reproductive power of χ^2 , we obtain

$$(3.11) \quad \begin{aligned} P\{N_0 < 1\} &= P\left\{\sigma_0 \sum_i \chi_{(p)i}^2 < 1 + \sum_{j=1}^{p-1} k_j \chi_{(p)j}^2\right\} \\ &= P\left\{\sigma_0 \chi_{(p)1}^2 < 1 + \sum_{j=1}^{p-1} k_j \chi_{(p)j}^2\right\}. \end{aligned}$$

By making special assumptions about the k_i , close approximations to the probability that N_D will be less than one, and hence that the process of iteration will converge, can be obtained. Instead of following this procedure, it is more desirable to have definite lower bounds for the probability that N_D will be less than one. This will lead to an overstatement in the number of decimal places of accuracy necessary in the first approximation C_0 to assure convergence, but it will practically eliminate the possibility of having to recalculate the first approximation, and hence will lead to greater efficiency in the long run.

4. The derivation of the formula for determining the required degree of accuracy. The inequality used in this section is derived in two steps from (3.10). Since $k_i > 0$ ($i = 1, \dots, p-1$) it follows immediately that

$$(4.1) \quad P\{N_D < 1\} > P\{\sigma_0 \chi_{(p^2)}^2 < 1\}.$$

In order to use this inequality, the upper bound for σ_0

$$(4.2) \quad \sigma_0 \leq (\text{tr } \sigma^t)^{1/t} \bullet$$

can be used. For $t = 1$,

$$(4.3) \quad \sigma_0 \leq \text{tr } \sigma = \text{tr } \mu A A' = \mu \text{tr } A A' = \mu N_A^2.$$

D1. Wald pointed out that using (4.2) for $t = 1$ reduces the amount of information retained in (4.1) to that which is contained in the inequality,

$$(4.4) \quad N(D) \leq N(A)N(E).$$

A closer upper bound is feasible in any particular case, and can be introduced at this point by letting $t = 2$ or $t = 3$. The following formula will be developed for the general case, making use of (4.3).

Substituting (4.3) in (4.1), we obtain

$$(4.5) \quad P\{N_D < 1\} > P\left\{\chi_{(p^2)}^2 < \frac{1}{\mu N_A^2}\right\}.$$

It is desirable to separate the effects of the order of the matrix A on convergence, and the order of magnitude of the elements. Hence we introduce as a measure of the average size of the a_i , their root mean square m , so that

$$(4.6) \quad m^2 = \sum_i \sum_j a_{ij}^2 / p^2$$

Hence

$$(4.7) \quad N_A = pm.$$

The final form of the inequality is

$$(4.8) \quad P\{N_D < 1\} < P\left\{\chi_{(p^2)}^2 < \frac{12 \cdot 10^{2k}}{p^2 m^2}\right\}.$$

First we will obtain an expression for the number of decimal places required in the first approximation to make the probability of convergence at least .999. Then the expression will be checked directly by means of (4.8) and tables of the chi-square function.

For large values of p , $\sqrt{2\chi_{(p^2)}^2}$ is approximately normally distributed with mean value $\sqrt{2p^2 - 1}$ and unit variance [2], [5]. Applying this transformation to the right hand side of (4.8), and noting that 3.1 standard deviations is slightly greater than the deviation corresponding to .999, we obtain as the condition for

$$(4.9) \quad P \left\{ \chi_{(p^2)}^2 < \frac{12 \cdot 10^{2k}}{p^2 m^2} \right\} \geq .999$$

or

$$(4.10) \quad P \left\{ 2\chi_{(p^2)}^2 < 2 \cdot \frac{12 \cdot 10^{2k}}{p^2 m^2} \right\} \geq .999$$

that it is sufficient that

$$(4.11) \quad \sqrt{\frac{24 \cdot 10^{2k}}{p^2 m^2}} - \sqrt{2p^2 - 1} \geq 3.1.$$

This is equivalent to

$$(4.12) \quad k \geq \log_{10} p + \log_{10} m + \log_{10} \left(\sqrt{p^2 - \frac{1}{2}} + \frac{3.1}{\sqrt{2}} \right) \\ - \frac{1}{2} \log_{10} 24 + \log_{10} \sqrt{2}.$$

Since the characteristic of a logarithm is insensitive to the argument, rounding off will introduce a negligible error, and we finally obtain an upper limit to the lower bound of k ,

$$(4.13) \quad k > \log_{10} m + \log_{10} p + \log_{10} (p + 3) - .55.$$

In order to verify the accuracy of (4.13) for small values of p , certain values of p , k and m are chosen and the probabilities associated with (4.8) determined [2]. The entries in brackets are the corresponding values of k determined from (4.13).

A typical example will illustrate the use of table on facing page. Let the matrix A to be inverted be a fourth order correlation matrix. The mean magnitude m is about $\frac{1}{2}$ and $p = 4$. If the first approximation C_0 is obtained to one place accuracy, then the probability that the sequence C_1, C_2, \dots will converge to A^{-1} will be greater than .999. Using formula (4.13), we obtain $k = .53$. Since one is the first integer greater than .53, the table verifies the use of the formula.

Although the formula was developed on the assumption that p is large, every value calculated is consistent with the table. This lends support to its use for small values of p .

*The Probability of Convergence of the Iterative Process**

	$\begin{matrix} p \\ k \end{matrix}$	2	3	4	5
$m = \frac{1}{2}$	-1	0+	0+	0+	0+
	0	[.05].982	[.33] 199	[.53]0+	[.70]0+
	1	1-	1-	1-	1-
$m = 2$	-1	0+	0+	0+	0+
	0	[.85].051	[.93]0+	0+	0+
	1	1-	1-	[1.13].715	[1.30]0-
	2	1-	1-	1-	1-
$m = 10$	0	0+	0+	0+	0+
	1	[1.35] 439	[1.63]0+	1[.83]0+	0+
	2	1-	1-	1-	[2.00]1-

* "1-" means greater than .999.

It has already been pointed out that k is not sensitive to rounding off of the argument of the logarithm. Thus for $p = 20$ and $m = 2$, we can let $\log_{10} m = 3$, $\log_{10} p = 1.3$, $\log_{10} (p + 3) = 1.36$ and obtain

$$k = .3 + 1.3 + 1.36 - .55 = 2.41,$$

from which it follows that three decimal place accuracy in C_0 will practically insure convergence of the iterative process.

5. The mean, variance, and exact distribution. To obtain the moments of N_D^2 , the most convenient form to use is (3.8). Since the $\chi_{(p),i}^2$ are independent

$$(5.1) \quad E[N_D^2] = E\left[\sum_i \sigma_i \chi_{(p),i}^2\right] = p \sum_i \sigma_i.$$

$$\begin{aligned}
 \sigma_{N_D^2} &= E[N_D^2] - (E[N_D^2])^2 \\
 &= E\left[\sum_{i=1}^p \sigma_i^2 (\chi_{(p),i}^2)^2 + 2 \sum_{i < j} \chi_{(p),i}^2 \chi_{(p),j}^2 \sigma_i \sigma_j\right] - (p \sum_i \sigma_i)^2 \\
 (5.2) \quad &= (2p + p^2) \sum_i \sigma_i^2 + 2p^2 \sum_{i < j} \sigma_i \sigma_j - p^2 \sum_i \sigma_i^2 - 2p^2 \sum_{i < j} \sigma_i \sigma_j \\
 &= 2p \sum_i \sigma_i^2
 \end{aligned}$$

These can be expressed in terms of the elements of A and the variance of the elements of E , since

$$\begin{aligned}
 \sum_i \sigma_i &= \text{tr}(\sigma) = \mu \text{tr}(AA') = \mu N_A^2, \\
 (5.3) \quad \sum_i \sigma_i^2 &= \text{tr} \sigma^2 = \mu^2 \text{tr}(AA'AA').
 \end{aligned}$$

The exact distribution of N_p^2 can be obtained readily when p is even. In this case the infinite integral,

$$(5.4) \quad \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-t N_p^2} dt}{(1 - i2\sigma_1 t)^{p/2} \dots (1 - i2\sigma_p t)^{p/2}},$$

can be evaluated by contour integration. The integral satisfies the conditions given in Whittaker and Watson [3, sec. 622], if the semicircle of the contour is taken on the lower half of the complex plane.

For the case $p = 2$, for example, there are simple poles, at $t = \frac{-i}{2\sigma_1}, \frac{-i}{2\sigma_2}$.

The sum of the residues at these poles, multiplied by i yields the exact distribution:

$$(5.5) \quad \frac{\sigma_1 e^{-N_p^2/2\sigma_1}}{2(\sigma_1 - \sigma_2)} + \frac{\sigma_2 e^{-N_p^2/2\sigma_2}}{2(\sigma_2 - \sigma_1)}.$$

For even values of p greater than 2, the values of the residues can be obtained by repeated differentiation.

6. Summary. We are now in a position to discuss the applicability of the results of this paper to the problem which arises most frequently in practice. The elements of the first approximation to the inverse will deviate from their expected values only because the first approximation is carried to a limited number of places, say k . In this case the deviations will be distributed with constant density over a range of length 10^{-k} . The elements of E , the matrix of deviations,

$$(6.1) \quad E = A^{-1} - C_0,$$

are now each independently distributed, but with uniform density, range 10^{-k} and mean equal to zero. From (2.4)

$$(6.2) \quad D = AE,$$

we observe that each element of D will be a linear combination of p independently and rectangularly distributed variables, each with mean zero and range 10^{-k} . The analysis of sections 3, 4, and 5 will be valid if d_{ij} can be considered to be normally distributed.

There is much experimental evidence and theoretical justification for assuming that the elements of D are normally distributed. A sufficient condition that the d_{ij} approach normality as p increases is that the sum of the a_{ij}^2 in any row of A be divergent as the order of the matrix approach infinity, while at the same time every element be less than some constant value independent of the order of the matrix [4].

The experimental and theoretical evidence supporting the approach of the d_{ij} to normality, the fact that the logarithms are insensitive to errors of approxi-

mation in their arguments and the fact that the lower bounds to the probability of convergence of the iterative process are used, all lend support to the formula

$$k > \log_{10} m + \log_{10} p + \log_{10} (p + 3) - .55$$

for determining the number of places (k) necessary in the first approximation (C_0) to the inverse of A , a matrix of order p whose elements have mean size m , to make the probability at least .999 that the process of iteration will yield a sequence of matrices which will converge to the true inverse. The ultimate justification of the use of this formula can only be by the results of its application in practice.

REFERENCES

- [1] HAROLD HOTELLING, "Some new methods in matrix calculation," *Annals of Math. Stat.*, Vol. 14 (1943) pp. 1-34.
- [2] FISHER and YATES, *Statistical Table for Biological, Agricultural and Medical Research*, p. 31.
- [3] WHITTAKER and WATSON, *Modern Analysis*, 4th edition, pp. 113-114
- [4] J. V. USPENSKY, *Introduction to Mathematical Probability*, 1935, Chapter 14
- [5] E. B. WILSON and M. M. HILFERTY, "The distribution of chi-square," *Proc. Nat. Acad. Science*, Vol. 17 (1931), pp. 684-688

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items

ON DISTRIBUTION-FREE TOLERANCE LIMITS IN RANDOM SAMPLING

BY HERBERT ROBBINS

Post Graduate School, Annapolis, Md.

Let X_1, \dots, X_n be independent random variables each with the continuous and differentiable cumulative distribution function $\sigma(x) = \Pr(X_i \leq x)$. A continuous function $f(x_1, \dots, x_n)$ with the property that the random variable $Y = \sigma(f(X_1, \dots, X_n))$ has a probability distribution which is independent of $\sigma(x)$ will be called a *distribution-free upper tolerance limit*¹ (d. f. u. t. l.). We shall prove

THEOREM 1. *A necessary and sufficient condition that the continuous function $f(x_1, \dots, x_n)$ be a d. f. u. t. l. is that the function*

$$f(x_1, \dots, x_n) = \prod_{i=1}^n \{f(x_1, \dots, x_n) - x_i\}$$

be identically zero.

PROOF. Since f is continuous, we can prove the necessity of the condition by deriving a contradiction from the assumption that f is a d. f. u. t. l. for which there exist distinct numbers a_1, \dots, a_n such that $f(a_1, \dots, a_n) = A \neq a_i$, ($i = 1, \dots, n$).

Since the numbers a_1, \dots, a_n, A are distinct, there will exist a positive number ϵ such that the $(n+1)$ intervals

$$I: \quad A - \epsilon \leq x \leq A + \epsilon$$

$$I_i: \quad a_i - \epsilon \leq x \leq a_i + \epsilon \quad (i = 1, \dots, n),$$

have no points in common. Moreover, since f is continuous, there will correspond to ϵ a positive number $\epsilon_1 < \epsilon$ such that

$$A - \epsilon \leq f(x_1, \dots, x_n) \leq A + \epsilon,$$

provided that simultaneously

$$|x_i - a_i| < \epsilon_1 \quad (i = 1, \dots, n).$$

Now let p be any number between $\frac{1}{2}$ and $\frac{3}{4}$. Corresponding to p we define the function $\sigma_p(x)$ as follows. In the interval I we set $\sigma_p(x) = p$. In every interval

$$J_i: \quad a_i - \epsilon_1 \leq x \leq a_i + \epsilon_1 \quad (i = 1, \dots, n)$$

¹ Cf. S. S. Wilks, *Mathematical Statistics*, Princeton University Press (1943), pp. 93-94.

we let $\sigma_p(x)$ increase an amount $\left(\frac{1}{3n}\right)$. Outside the intervals I, J_1, \dots, J_n we define $\sigma_p(x)$ in any manner so that it is continuous, differentiable, and non-decreasing for every x , and has the properties $\sigma_p(-\infty) = 0, \sigma_p(\infty) = 1$. It is clear that we can do this

Let S denote the set of all points (x_1, \dots, x_n) of n dimensional space such that simultaneously

$$|x_i - a_i| \leq \epsilon_1 \quad (i = 1, \dots, n).$$

Then by construction, for $\sigma_p(x)$ defined above,

$$Pr((X_1, \dots, X_n) \in S) = \left(\frac{1}{3n}\right)^n.$$

But if $(X_1, \dots, X_n) \in S$, then by construction,

$$A - \epsilon \leq f(X_1, \dots, X_n) \leq A + \epsilon$$

and

$$Y = \sigma_p(f(X_1, \dots, X_n)) = p.$$

Hence for $\sigma(x) = \sigma_p(x)$ we have

$$Pr(Y = p) \geq \left(\frac{1}{3n}\right)^n.$$

But since f is a d. f. u. t. l., this inequality must hold for any $\sigma(x)$

Now choose a set of numbers

$$\frac{1}{3} < p_1 < p_2 < \dots < p_m < \frac{2}{3},$$

where $m = 2(3n)^n$. Then from the above,

$$Pr(Y = \text{one of the numbers } p_1, \dots, p_m) \geq 2.$$

This is the desired contradiction.

Let $O_r(x_1, \dots, x_n)$ be the function whose value is the r th term when the numbers x_1, \dots, x_n are arranged in non-decreasing order of magnitude. In terms of the functions O_r we can characterize the continuous functions f which satisfy the identity $\dot{f} \equiv 0$ as follows. Let i_1, \dots, i_n be a permutation of the integers $1, \dots, n$. Denote by $E(i_1, \dots, i_n)$ the set of all points (x_1, \dots, x_n) such that

$$x_{i_1} < x_{i_2} < \dots < x_{i_n}.$$

The $n!$ sets E are open and disjoint. Since f is continuous and $\dot{f} \equiv 0$, in each $E(i_1, \dots, i_n)$ we must have, for some r ,

$$f(x_1, \dots, x_n) \equiv O_r(x_1, \dots, x_n),$$

where the integer $r = r(i_1, \dots, i_n)$ must depend on the permutation i_1, \dots, i_n in such a way that f may be extended continuously over the whole space. (The

condition for this is as follows. Two permutations i_1, \dots, i_n and j_1, \dots, j_n may be called *adjacent* if they differ only by an interchange of two adjacent integers. Then for any two adjacent permutations, either $r(i_1, \dots, i_n) = r(j_1, \dots, j_n)$ or the two values of r are the two interchanged integers. For example, the function

$$f(x_1, x_2, x_3) = \begin{cases} O_3(x_1, x_2, x_3) & \text{if } O_3(x_1, x_2, x_3) = x_1 \\ O_2(x_1, x_2, x_3) & \text{otherwise} \end{cases}$$

satisfies this requirement.)

We shall now prove that the necessary condition, $\hat{f} = 0$, of Theorem 1 is sufficient to ensure that the continuous function f be a d. f. u. t. l. From the argument of the preceding paragraph, any continuous function f such that $\hat{f} \equiv 0$ will in each set $E(i_1, \dots, i_n)$ have the value $O_r(x_1, \dots, x_n)$, where r is an integer from 1 to n . Since the variables X_1, \dots, X_n are independent and have the same probability distribution, the probability that (X_1, \dots, X_n) will belong to $E(i_1, \dots, i_n)$ is equal to $(1/n!)$ for every permutation i_1, \dots, i_n . Let

$$W = f(X_1, \dots, X_n).$$

Then if $\varphi(x) = d\sigma(x)/dx$ denotes the probability density function of each X_i , the conditional p. d. f. of $W = O_r(X_1, \dots, X_n)$, given that (X_1, \dots, X_n) belongs to $E(i_1, \dots, i_n)$, will be $n!\psi_r(w)$, where

$$\psi_r(w) = \frac{\varphi(w)\sigma^{r-1}(w)[1 - \sigma(w)]^{n-r}}{(r-1)!(n-r)!}.$$

Thus $\psi_r(w)$ will be of the form

$$\psi_r(w) = \varphi(w)F_r(\sigma(w)),$$

where $F_r(\sigma(w))$ is a polynomial in $\sigma(w)$. Hence the conditional p. d. f. of $Y = \sigma(W)$, given that (X_1, \dots, X_n) belongs to $E(i_1, \dots, i_n)$, will be $n!\xi_r(y)$, where

$$\xi_r(y) = F_r(y),$$

and the p. d. f. of Y will be

$$\xi(y) = \sum F_r(y),$$

where the summation is over the $n!$ integers $r = r(i_1, \dots, i_n)$. This is independent of $\sigma(x)$, so that f is a d. f. u. t. l. This completes the proof of Theorem 1.

A function $f(x_1, \dots, x_n)$ is *symmetric* if its value is unchanged by any permutation of its arguments. It is clear that the only continuous and symmetric functions f which satisfy the identity $\hat{f} \equiv 0$ are the n functions $O_r(x_1, \dots, x_n)$. Hence we can state

THEOREM 2. *The only symmetric d. f. u. t. l.'s are the n functions $O_r(x_1, \dots, x_n)$ ($r = 1, \dots, n$).*

A FORMULA FOR SAMPLE SIZES FOR POPULATION TOLERANCE LIMITS

BY H. SCHEFFÉ AND J. W. TUKEY

Princeton University

In a paper to appear in a later issue of this journal dealing with various results on non-parametric estimation, we shall discuss in detail an approximate formula for the numerical calculation of sample sizes for Wilks' population tolerance limits. Because of the practical usefulness of this formula, it seems desirable to make it available without delay. Its accuracy is adequate for all direct applications.

An interval I is said to cover a proportion π of a univariate population with cumulative distribution function $F(x)$ if $\int_I dF = \pi$. Let X_1, X_2, \dots, X_n be a random sample from the population, and $Z_1 \leq Z_2 \leq \dots \leq Z_n$ be a rearrangement of X_1, X_2, \dots, X_n . Define $Z_0 = -\infty, Z_{n+1} = +\infty$, and consider the proportion B of the population covered by the random interval (Z_k, Z_{n-m+1}) . Then $\Pr\{B \geq b\}$ is independent of $F(x)$ if $F(x)$ is continuous¹, and equals $1 - I_b(n - r + 1, r)$, where $r = k + m$ and $I_x(p, q)$ is K. Pearson's notation for the incomplete Beta function.

Choose a confidence coefficient $1 - \alpha$, a pair of positive integers k, m , and a fraction b . The sample size n for which we can make the statement "the probability is $1 - \alpha$ that the random interval (Z_k, Z_{n-m+1}) cover a proportion b or more of the population" is then determined by the equation

$$(1.1) \quad I_b(n - r + 1, r) = \alpha,$$

where $r = k + m$. Our approximate solution is

$$(1.2) \quad n = \frac{1}{4}\chi_\alpha^2(1 + b)/(1 - b) + \frac{1}{2}(r - 1),$$

where χ_α^2 is the 100 α percent point on the χ^2 -distribution with $2r$ degrees of freedom. The required values of χ_α^2 may be found for $\alpha = .1, .05, .025, .01, .005$ in Catherine Thompson's table [2]. For this range of α , and for $b \geq .9$, extensive numerical calculations indicate that the error of (1.2) is less than one tenth of one percent, and is always positive, that is, n is slightly overestimated by (1.2). We have not yet obtained an analytic proof of this statement, which refers to the difference from the exact (and, in general, non-integral) solution of (1.1).

As explained elsewhere [1], formula (1.2) may be used for Wald's solution of the multivariate case.

REFERENCES

- [1] H. SCHEFFÉ, *Annals of Math. Stat.*, Vol. 14 (1943), p. 324
- [2] C. M. THOMPSON, *Biometrika*, Vol. 32 (1941), p. 189.

¹ That the theory is valid in this case we show later. Previous proofs have required the continuity of $F'(x)$

A GENERALIZATION OF WARING'S FORMULA

BY T. N. E. GREVILLE

Bureau of the Census

Waring's formula (frequently, but less correctly, called Lagrange's formula) gives the polynomial of degree n taking on specified values for $n + 1$ distinct arguments. It is frequently used for interpolation purposes in dealing with functions for which numerical values are given at unequal intervals. This formula may be written in the form:

$$(1) \quad f(x) = \sum_{i=0}^n \left[f(a_i) \prod_{j \neq i} \frac{x - a_j}{a_i - a_j} \right],$$

where $a_0, a_1, a_2, \dots, a_n$ are the arguments for which the value of the polynomial $f(a)$ is given. This formula was first published by Waring [2] in 1779, and it was not until 1795 that Lagrange gave it in his book: *Leçons Élémentaires sur les Mathématiques*. The prominent British actuary and mathematician, Mr. D. C. Fraser states that "there are identities of notation in the statement of the formula which leave little doubt that Lagrange was simply quoting from Waring's paper." Waring's priority was brought to my attention by Mr. Fraser and by Dr. W. Edwards Deming.

If any two or more of the arguments a_i are equal, the form (1) becomes indeterminate. However, the limiting value, as $m + 1$ specified arguments approach a common value a , can be shown to be an expression involving the first m derivatives of the polynomial $f(x)$ for the argument a . This case of "repeated arguments" is of considerable interest, especially in connection with the theory of osculatory, or smooth-junction interpolation [1, p. 33]. It is the purpose of this note to generalize the formula (1) to the case in which not only the value of $f(x)$ but also of its first m_i derivatives are given for each argument a_i . The degree of the polynomial represented, which we shall denote by N , is $n + \sum_{i=0}^n m_i$.

The generalized formula is:

$$(2) \quad f(x) = \sum_{i=0}^n \left[P_i(x - a_i) \prod_{j \neq i} \frac{(x - a_j)^{m_j+1}}{(a_i - a_j)^{m_j+1}} \right],$$

where $P_i(x - a_i)$ denotes a polynomial in $x - a_i$ obtained by the following procedure. First, $f(x)$ is expanded in a Taylor series in powers of $x - a_i$. Next, the expression $\left(1 + \frac{x - a_i}{a_i - a_j}\right)^{m_j+1}$ is expanded as a binomial series for every j different from i . Finally, all the $n + 1$ expansions (n binomial and one Taylor) are multiplied together, and all terms containing powers of $x - a_i$ higher than m_i are rejected. This formula has already been given by Steffensen [1, p. 33] for the particular case in which every $m_i = 1$.

The general formula (2) is difficult to arrive at without a previous knowledge

of the result, but is easily shown to be the correct expression. Upon differentiating k times ($0 \leq k \leq m_r$) all the terms in the summation except the one corresponding to $i = r$ will contain the factor $(x - a_r)^{m_r - k + 1}$, and will therefore vanish for $x = a_r$. Moreover, the non-vanishing term, before differentiation, will agree, up to and including terms containing $(x - a_r)^{m_r}$, with the Taylor expansion of $f(x)$ in powers of $x - a_r$, since the product expression within the brackets will be exactly canceled, as far as terms of degree m_r , by the n binomial expansions. Hence the k th derivative of the non-vanishing term in the summation will be $f^{(k)}(a_r)$ for $x = a_r$. This establishes the formula.

This formula is clearly equivalent to the Newton divided difference interpolation formula with repeated arguments [1, p. 33], the argument a_i occurring $m_i + 1$ times. Therefore, if $f(x)$ is any function other than a polynomial of degree N or less, it is necessary to add a remainder term [1, pp. 22-23] of the form

$$f_N(x) \prod_{i=0}^n (x - a_i)^{m_i+1},$$

where $f_N(x)$ denotes the limiting value [1, pp. 20-21] of the divided difference of order N involving the arguments x, a_0, a_1, \dots, a_n , with each argument a_i appearing $m_i + 1$ times. The existence of all the indicated derivatives is, of course, essential.

REFERENCES

- [1] J. F. Steffensen, *Interpolation*, Baltimore, 1927
- [2] E. Waring, "Problems concerning interpolation," *Phil. Trans. Royal Soc.*, Vol. 69 (1779), pp. 59-67

NOTE ON THE VARIANCE AND BEST ESTIMATES

BY H. G. LANDAU

Washington, D. C.

The purpose of this note is to point out a certain relation between the variances, σ_1^2 and σ_2^2 , of the random variables, x_1 and x_2 , and the probabilities,

$$P_1(t) = \Pr[|x_1 - E(x_1)| < t]$$

$$P_2(t) = \Pr[|x_2 - E(x_2)| < t].$$

This is, if $\sigma_1^2 < \sigma_2^2$, then $P_1(t) > P_2(t)$ in at least one interval, $t_1 < t < t_2$.

A note by A. T. Craig [1] gave an example for which it was stated that $\sigma_1^2 < \sigma_2^2$ and $P_1(t) \leq P_2(t)$ for every t ; but, as was pointed out by Neyman [2], calculation of the probabilities involved shows the statement to be incorrect.

The present result provides a certain justification for the use of minimum variance estimates by assuring that no other estimate with the same mean can have, for every value of t , a greater probability of a deviation from the mean

less than t . If an estimate can be found which has a greater value of $P(t)$ for all t than does any other estimate, it is necessarily the minimum variance estimate.

The theorem below includes a similar relation for equal variances. This theorem can be obtained from known general results on inequalities for distributions determined by moments, [3] and [4]. The formulation given here with its significance for estimates does not appear to have been remarked.

THEOREM. *If the random variables, x_1 and x_2 , have finite variances, σ_1^2 and σ_2^2 , and*

$$\sigma_1^2 \leq \sigma_2^2,$$

then, either

$$Q(t) = P_1(t) - P_2(t),$$

is equal to zero at all points of continuity, which can occur only for $\sigma_1^2 = \sigma_2^2$, or there is an interval, $t_1 < t < t_2$, in which $Q(t)$ is positive.

PROOF. We write the variance as the Stieltjes integral,

$$\sigma_1^2 = \int_0^\infty t^2 dP_1(t),$$

and similarly for σ_2^2 .

Let

$$\begin{aligned} S(T) &= \int_0^T t^2 dP_1(t) - \int_0^T t^2 dP_2(t) = \int_0^T t^2 dQ(t) \\ &= T^2 Q(T) - 2 \int_0^T t Q(t) dt, \end{aligned}$$

integrating by parts.

Now

$$T^2[1 - P_1(T)] = T^2 \int_T^\infty dP_1(t) \leq \int_T^\infty t^2 dP_1(t),$$

and since σ_1^2 is finite, $\int_T^\infty t^2 dP_1(t) \rightarrow 0$ as $T \rightarrow \infty$, so that $\lim_{T \rightarrow \infty} T^2[1 - P_1(T)] = 0$, and similarly for $P_2(t)$.

Hence $T^2 Q(T) = T^2[1 - P_2(T)] - T^2[1 - P_1(T)] \rightarrow 0$ as $T \rightarrow \infty$, and since by definition $\lim_{T \rightarrow \infty} S(T) = \sigma_1^2 - \sigma_2^2$ it follows that

$$\sigma_1^2 - \sigma_2^2 = -2 \int_0^\infty t Q(t) dt.$$

From this it can be seen that either, $Q(t)$ vanishes at all points of continuity, in which case $\sigma_1^2 = \sigma_2^2$, or $Q(t)$ must be positive in some interval, since otherwise $\int_0^\infty t Q(t) dt$ must be negative and hence $\sigma_1^2 - \sigma_2^2 > 0$ contrary to the assumption, $\sigma_1^2 \leq \sigma_2^2$.

REFERENCES

- [1] A. T. CRAIG, "A note on the best linear estimate," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 88-90.
- [2] J. NEYMAN, *Math. Reviews*, Vol. 4 (1943), p. 280
- [3] J. V. USPENSKY, *Introduction to Mathematical Probability*, New York, McGraw-Hill (1937), pp. 373-380.
- [4] A. WALD, "Limits of a distribution function determined by absolute moments and inequalities satisfied by absolute moments," *Trans. Amer. Math. Soc.*, Vol. 46 (1939), pp. 280-306.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of general interest

Personal Items

Dr. R. L. Anderson, on leave from the North Carolina State College, is serving as a research mathematician on a war research project at Princeton University.

Professor Kenneth J. Arnold, on leave from the University of New Hampshire, is doing war research in the Statistical Research Group, at Columbia University.

Dr. W. J. Dixon, on leave from the University of Oklahoma, is serving as a research mathematician on a war research project at Princeton University.

Mr. R. M. Foster of Bell Telephone Laboratories has been appointed professor and head of the department of mathematics of the Polytechnic Institute of Brooklyn.

Dr. Hilda Geringer, Lecturer at Bryn Mawr College, has been appointed Professor of Mathematics and Head of the Mathematics Department at Wheaton College, Norton, Massachusetts.

Assistant Professor E. H. C. Hildebrandt of the State Teachers College, Upper Montclair, New Jersey, has been appointed to an assistant professorship at Northwestern University.

Mr. John F. Kenney of the University of Wisconsin has been promoted to an assistant professorship.

Assistant Professor L. A. Knowler of the University of Iowa has been promoted to an associate professorship.

Dr. Saul B. Sells is now Head Statistician of the Statistical Standards Branch, Office of Price Administration, Washington.

Mr. Walter H. Thompson, formerly an instructor at Virginia Polytechnic Institute, is now associated with the Kellogg Corporation, New York.

Professor Helen M. Walker of Teachers College, Columbia University, has been elected president of the American Statistical Association.

Professor C. C. Wagner of Pennsylvania State College has been appointed assistant dean of the School of Liberal Arts.

Dr. Edward Helly of the Illinois Institute of Technology died November 28, 1943.

New Members

The following persons have been elected to membership in the Institute:

Andrews, T. Gaylord. Ph.D. (Nebraska) Instructor in Psychology, Barnard College, Columbia University, New York City.

Angell, Dorothy T. Member of Technical Staff, Bell Telephone Laboratories. *Murray Hill, New Jersey.*

Barnes, John L. Member of Technical Staff, Bell Telephone Laboratories. *21 North Cherry Lane, Rumson, New Jersey.*

- Bloom, Rose. B A (Hunter) Pvt, Billings General Hospital, Ft Benj Harrison, Indiana
- Bonnar, Robert Underwood. M.S (Univ of Washington) Associate Chemist, Bureau of Ships, Navy Dept *414 Whitestone Road, Silver Spring, Maryland.*
- Brearty, C. R. B S (California) Officer in Charge, Quality Control Section, Signal Corps Inspection Agency *19 West 4th St, Dayton 2, Ohio*
- Campbell, George Clyde. M S. (Iowa) Captain CWS Chief, Fiscal Div., Pine Bluff Arsenal. *Troy Road, RFD #1, Boonton, New Jersey.*
- Clifford, Paul C. A M (Columbia) Asst Prof of Math., State Teachers College, Montclair, New Jersey. *541 Upper Mountain Ave*
- Cobb, William J Statistician, Census Bureau *4036 8th St, NE, Washington, D. C.*
- Coggins, Paul Pond. M A (Harvard) Accountant, American Tel and Tel *195 Broadway, New York, N. Y.*
- Dietzold, Robert L. Ph.B. (Yale) M T S, Bell Telephone Lab. *34 W 11th St, New York, N Y*
- Elconin, Victor. M S (California Inst of Technology) Associate Physicist, California Inst of Technology *740 Cordova Ave, Glendale 6, California*
- Ferrell, Enoch B. M.A. (Oklahoma) Member of Technical Staff, Bell Telephone Laboratories. *75 Fuller Ave., Chatham, New Jersey*
- Ferris, Charles Duncan. A.B (Princeton) Engineering Statistician, Surveillance Branch, T/3 Army. Ballistic Research Laboratory, Aberdeen Proving Ground, Md
- Goldberg, Henry. M A (Columbia) Asst Mathematical Statistician, Columbia University, *401 West 118th Street, New York 27, N Y*
- Gordon, Donald A. A M. (Columbia) Assistant, Columbia University *1327 E 26 Street, Brooklyn, New York.*
- Greenleaf, Herrick E. H. Ph D. (Indiana) Prof of Mathematics *1024 S. College Avenue, Greencastle, Indiana*
- Griffitts, C H. Ph D (Michigan) Professor of Psychology *1507 Charlton Ave, Ann Arbor, Michigan*
- Hadley, Clausin D. Ph D. (Wisconsin) Statistician, Marketing Research Dept., Eli Lilly and Company, Indianapolis 6, Indiana
- Halbert, K. W. A M (Harvard) Statistician, Amer Tel. and Tel Co, *195 Broadway, New York, N Y*
- Hall, Marguerite F. Ph D (Michigan) Asst Prof. Public Health Statistics *25 Ridgeway, Ann Arbor, Michigan*
- Halmos, Paul Richard. Ph D (Illinois) Asst. Professor of Mathematics, Syracuse University *513 Fellows Avenue, Syracuse 10, N. Y*
- Harold, Miriam S. B A (Hunter) Member of Technical Staff, Bell Telephone Laboratories *19 Hillside Avenue, Chatham, New Jersey*
- Hatke, Sister M. Agnes. M S (Indiana State Teachers College) Graduate student at Purdue University. *St. Francis College, Lafayette, Indiana*
- Hizon, Manuel O. M A (Michigan) Philippine Govt Scholar at Univ of Michigan *816 Packard, Ann Arbor, Michigan*
- Hodgkinson, William, Jr. B.A. (Harvard) Major, HQ AAF, U S Army *195 Broadway, New York, N Y*
- Jacob, Walter C. Ph D (Cornell) Lt., U S Navy Bureau of Ships, Navy Dept., Washington, D C.
- Jones, Howard L. C.P A. (Illinois) Supervisor of Revenue Results, Illinois Bell Telephone Co *Room 1100, 309 W Washington St., Chicago 6, Ill.*
- Kaitz, Hyman B. A.B. (George Washington Univ.) Statistical Consultant, Psychological Research Unit No. 11, B A.A.F., Fort Myers, Fla.
- Keyfitz, Nathan. B Sc. (McGill) Statistician, Dominion Bureau of Statistics, Ottawa, Canada *Billings Bridge, Ontario.*
- Kirchen, Calvin J. M A (Wisconsin) Product Control Statistician, U S Rubber Co *4028 11th St Place, Des Moines 13, Iowa*

- LaSala, Lucy Anne.** B.A. (Hunter) Applied Math Group, Columbia Univ. 256 Irving Avenue, Brooklyn 27, New York.
- Leone, Fred Charles.** M.S. (Georgetown Univ.) Instructor in Math, Purdue Univ. 310 N. Salisbury St., W. Lafayette, Indiana
- McNamara, Kathryn J.** M.A. (Clark Univ.) Economist, Business Research Dept., H. J. Heinz Company, P O Box 57, Pittsburgh 30, Pa.
- McPherson, John Cloud.** B.S. (Princeton) Director of Engineering, Int'l Business Machines Corp 590 Madison Ave., New York 22, N. Y.
- Millikan, Max F.** Ph.D. (Yale) Asst. Director, Div. of Ship Reqs., War Shipping Admin. 2313 Huidkoper Pl., NW, Washington 7, D. C.
- Morrow, Dorothy Jeanne.** M.S. (Univ of Washington) Fellow in Mathematical Statistics, Columbia University. 605 W 115 St., New York 25, N. Y.
- Morse, John W.** M.A. (Columbia) Ordnance Engineer, Quality Control Unit Inspection Sec, Ammunition Branch, War Dept 11 Verne St., Bethesda, Md
- Mottley, Charles McCammon.** Ph.D. (Toronto) Lieut, USNR, Bureau of Ships, (333), Navy Dept, Washington, D. C.
- Murphy, Ray B.** B.A. (Princeton) 2nd Lt., U.S. Marine Corps Reserve. 28 Galfrey Rd., Upper Montclair, N. J.
- Myslivec, Vaclav.** Ph.D. (Prague) Czechoslovak delegate to the United Nations Interim Commission on Food and Agriculture Room 606, 1775 Broadway, New York 19, N. Y.
- Nicholson, George Edward, Jr.** M.A. (Univ. of North Carolina) Asst. Mathematician, Applied Mathematics Group, Columbia Univ. 176 Park St., Montclair, N. J
- Osterman, Herbert William.** B.S. (Michigan) c/o B. E. Wyatt, 1029 Vermont Ave. NW, Washington, D. C
- Parke, Nathan Grier, III.** A.B. (Princeton) Senior Aviation Design Research Engr, Navy Dept Malvern Ave., Ruxton 4, Md.
- Priestley, Alice E. A.** M.A. (New York Univ) Instructor in Math, Lafayette College. 228 McCartney St., Easton, Pa
- Rapkin, Chester.** M.A. (American University) Associate Statistical Analyst, Deputy Director, Division of Operating Statistics, Federal Home Loan Bank Admin., 2 Park Avenue, New York 16, N. Y.
- Rivoli, Bianca.** M.A. (Columbia University) Statistician, Judson Health Center. 3265 Bainbridge Avenue, New York 67, N. Y.
- Roshal, Sol M.** B.S. (Chicago) Officer in charge of research statistics, PRU #1, Nashville Army Air Center, Nashville, Tenn.
- Ross, Frank A.** Ph.D. (Columbia) Editor, Journal of the American Statistical Assn. Thelford, Vermont.
- Schaeffer, Esther.** A.B. (Chicago) Technical Asst., Univ. of Michigan. 547 Elm Street, Ann Arbor, Mich
- Schilling, Walter.** M.D. (Harvard) Asst. Clinical Professor of Medicine, Stanford University Hospital, San Francisco 15, California.
- Shannon, Claude E.** Ph.D. (M.I.T.) Member of Technical Staff, Bell Telephone Laboratories, 463 West Street, New York, N. Y.
- Sherman, Jack.** Ph.D. (Calif Inst. of Technology) Research Chemist, The Texas Co. 170 Church St., Poughkeepsie, New York.
- Smallwood, Hugh M.** Ph.D. (Johns Hopkins) Asst. Dept. Head, Physical Research Dept., U. S. Rubber Co., Market & South Streets, Passaic, N. J.
- Smith, Joan Thiede.** B.S. (Minnesota) Accountant. 673 East Nebraska Ave., St. Paul, Minn.
- Smith, R. Tynes, III.** Chief, Transport Economics Branch, Traffic Control Div., Office Chief of Transportation, A. S. F 1001 16th St. S., Arlington, Va.
- Sobczyk, Andrew.** Ph.D. (Princeton) Staff Member, Radiation Lab., Mass. Inst. of Technology 32 Bow Street, Lexington 72, Mass.

- Thurstone, Louis Leon.** Ph.D. (Chicago) Professor of Psychology, University of Chicago, Chicago, Illinois.
- Toralballe, Leopoldo V.** Ph D. (Michigan) Special Instructor, University of Michigan. *1109 Willard St , Ann Arbor, Mich*
- Walsh, John E.** B S (Notre Dame) Mathematician, Lockheed Aircraft Corp *707 East Elk Ave , Glendale 5, Calif*
- Weber, Bruce Travis.** M.A. (Columbia) Member of Technical Staff, Bell Telephone Laboratories, Marray Hill, New Jersey.
- Weiner, Louis.** A M (Harvard) Economist, U S. Bureau of Labor Statistics. *4215 Russell Ave., Mt. Rainier, Maryland.*
- Woodward, Patricia.** Ph.D (Pennsylvania) Associate Executive Secretary, Committee on Food Habits, National Research Council. *2101 Constitution Ave , Washington 25, D. C.*
-

REPORT ON THE WASHINGTON MEETING OF THE INSTITUTE

The second regional spring meeting of the Institute of Mathematical Statistics was held at George Washington University, Washington, D. C., Saturday and Sunday, May 6 and 7, 1944, jointly with a regional meeting of the American Statistical Association. The 383 registrants at the joint meeting included the following 76 members of the Institute:

Paul H. Anderson, R. L. Anderson, Theodore W. Anderson, Jr., Kenneth J. Arnold, Kenneth J. Arrow, 1st Lt., AC, Blair M. Bennett, Ernest S. Blanche, C. I. Bliss, Bonnar Brown, Joseph G. Bryan, Marjorie F. Buck, A. George Carlton, C. W. Churchman, William J. Cobb, Major A. C. Cohen, Jr., Edwin L. Crosby, Haskell B. Curry, J. H. Curtiss, R. B. Day, Scott Dayton, W. Edwards Deming, Philip Desmond, W. J. Dixon, J. L. Doob, Will Feller, William C. Flaherty, Thomas N. E. Greyville, E. J. Gumbel, Trygve Haavelmo, Margaret Jarman Haggood, Morris H. Hansen, Elyn A. Hoy, Leonard Hurwicz, Lt. (jg) W. C. Jacob, Alice S. Kautz, Evelyn M. Kennedy, Lila F. Knudsen, H. S. Konijn, T. Koopmans, Anita R. Kury, Jacob E. Lieberman, Philip J. McCarthy, Francis McIntyre, William G. Meadow, Lt. C. J. Maloney, Sophie Marcuse, John W. Mauchly, A. M. Mood, Vladimir A. Nekrassoff, Monroe L. Norden, H. W. Norton, Victor Perlo, A. C. Rosander, William Salkind, Marion M. Sandomire, Max Sasuly, Franklin E. Satterthwaite, S. B. Sells, L. W. Shaw, W. Arthur Shelton, Walter A. Shewhart, Harry Shulman, Blanche Skalak, John H. Smith, R. T. Smith, Arthur Stein, Joseph Steinberg, J. W. Tukey, Joseph L. Ullman, David F. Votaw, Jr., Capt. A. N. Watson, Frank M. Weida, Louis Weiner, S. S. Wilks, Patricia Woodward, Bertram Yood

All sessions were held jointly with the American Statistical Association.

Professor Will Feller of Brown University acted as Chairman for the Saturday afternoon session. The following papers were presented:

- 1 *Elements of the Theory of Testing Hypotheses.*
J. H. Curtiss, Jr., Navy Department
- 2 *Large Sample Tests of Statistical Hypotheses.*
Abraham Wald, Columbia University

Professor Frank Weida of George Washington University acted as Chairman for the Sunday morning session. The following contributed papers were presented:

- 1 *On the Statistics of Sensitivity Data.*
C. West Churchman and Benjamin Epstein, Frankford Arsenal and the University of Pennsylvania
- 2 *Simplified Plotting of Statistical Observations.*
E. J. Gumbel, New School of Social Research
- 3 *Distribution of Sample Variances and Covariances of Normally Distributed Noncentral Variables.*
M. A. Girshick, Department of Agriculture
- 4 *An Application of the Variate Difference Method to Multiple Regression.*
Gerhard Tintner, Department of Agriculture
- 5 *Autocorrelation in London Temperature.*
Horace W. Norton, Department of Commerce

Professor Samuel S. Wilks of Princeton University acted as Chairman for the Sunday afternoon session. The following papers were presented:

- 1 *Regression Problems in Time Series*
Tjalling Koopmans, Combined Shipping Adjustment Board
- 2 *Foundations of the Theory of Time Series.*
J. L. Doob, Navy Department

The business meeting of the Washington Chapter of the Institute was held Sunday morning. The proposed Constitution of the Washington Chapter was ratified, and elections under this constitution were held for the first time. The following officers were elected:

William G. Madow, Census Bureau—3 year term, Secretary of the Program Committee
1944-45

Solomon Kullback, War Department—2 year term

Frank Weida, George Washington University—1 year term.

W. G. MADOW
Secretary, Program Committee
Washington Chapter

THE ELEMENTARY GAUSSIAN PROCESSES

By J. L. DOOB

University of Illinois

1. Introduction

One of the simplest interesting classes of temporally homogeneous stochastic processes is that for which the distributions of the defining chance variables $\{x(t)\}$ are Gaussian. It is supposed that

(A) if $t_1 < \dots < t_r$, the multivariate distribution of $x(t_1), \dots, x(t_r)$ is Gaussian,¹ and that

(B) this distribution is unchanged by translations of the t -axis.

The process is N -dimensional if $x(t)$ is an N -tuple $x_1(t), \dots, x_N(t)$. The means $E\{x(t)\}^2$ are independent of t , and will always be supposed to vanish in the following discussion.

The correlation matrix function $R(t): (r_{ij}(t))$ is defined by

$$(1.1.1) \quad r_{ij}(t) = E\{x_i(s)x_j(s+t)\}.$$

This expectation is independent of s , because of condition (B). The matrix function $R(t)$ satisfies the equation

$$(1.1.2) \quad r_{ij}(t) = r_{ji}(-t), \quad i, j = 1, \dots, N.$$

It follows that when $t = 0$ the matrix is symmetric:

$$(1.1.3) \quad r_{ij}(0) = r_{ji}(0), \quad i, j = 1, \dots, N,$$

and it is also well known that $R(0)$ is non-negative definite. Conditions on the functions $r_{ij}(t)$ necessary and sufficient that $R(t)$ be the correlation matrix function of a stochastic process were found for the case $N = 1$ by Khintchine³ and for all N by Cramér.⁴

Hypothesis (A), that the process is Gaussian seems at first a restriction so strong that Gaussian processes are unimportant. These processes are, however, of fundamental importance, for the following reasons

(i) If $R(t)$ is the correlation matrix function of *any* temporally homogeneous stochastic process, there is, according to Khintchine and Cramér, a Gaussian process with this same correlation function. This Gaussian process is uniquely determined by the correlation function (assuming that all first order moments vanish, as usual). Because of this intimate connection between the temporally homogeneous Gaussian processes and the most general temporally homogeneous

¹ Singular Gaussian distributions will not be excluded. For example the $x(t_i)$ may all vanish identically.

² The expectation of a chance variable x will be denoted by $E\{x\}$

³ *Mathematische Annalen*, Vol 109 (1934), p. 608.

⁴ *Annals of Math*, Vol 41 (1940), pp 215-230.

processes, it is not surprising that very few facts are known about specifically Gaussian processes, that is facts which are true of temporally homogeneous Gaussian processes, but not of temporally homogeneous processes in general.

(ii) It follows from (i) that in any investigation of temporally homogeneous stochastic processes involving only first and second moments—for example least squares prediction by linear extrapolation—it may be assumed that the variables are Gaussian. Under this hypothesis, the investigator may be helped by the suggestive specialized interpretations possible in the Gaussian case of results which hold in the general case. For example if $N = 1$, the least squares best prediction in the Gaussian case for $x(n+1)$ in terms of a linear combination of the variables $x(1), \dots, x(n)$ is the conditional expectation of $x(n+1)$ for given $x(1), \dots, x(n)$, which is the least squares best prediction of $x(n+1)$ in terms of $x(1), \dots, x(n)$ with no restriction on the functions involved. Thus the linearity of the prediction, which must be part of the hypothesis in the general case, is automatically true in the Gaussian case. There is necessarily a linear least squares best prediction of $x(n+1)$ in terms of the complete past $\dots, x(n-1), x(n)$ since the corresponding conditional expectation is certainly defined in the Gaussian case, and is linear in that case.

(iii) In many applications, there is a real justification for hypothesis (A) that the process is Gaussian. This is so in certain physical studies, for example, because the Maxwell distribution of molecular velocities is Gaussian. Examples will be given below.

The processes discussed in the present paper are all temporally homogeneous Gaussian processes. Most of the theorems will be valid for any temporally homogeneous processes for which the second moments of the variables exist,⁵ with the following changes: independent chance variables which are linear combinations of the $x(s)$ will become merely uncorrelated chance variables; the convergence with probability 1 of a series of such chance variables will become merely convergence in the mean; the conditional expectation of one such variable y for given values of others, y_1, y_2, \dots will become merely the linear approximation $\sum_j a_j y_j$ of y in terms of the y_j which minimizes

$$E\{[y - \sum_j a_j y_j]^2\},$$

that is to say the conditional expectation becomes the least squares linear approximation.

The following theorem and its corollary are fundamental in the study of linear prediction involving infinitely many variables. The results are implicit in much of the work on the subject but do not seem to have been stated explicitly before.

THEOREM 1.2. *Let \dots, x_0, x_1, \dots be a sequence of one-dimensional Gaussian*

⁵ The processes need not even be temporally homogeneous. It is necessary only that $E\{x(s)\}$ and $E\{x(s)x(s+t)\}$ be independent of s .

chance variables with the property that if $n_1 < \dots < n_r$, the multivariate distribution of x_{n_1}, \dots, x_{n_r} is Gaussian and that

$$(1.2.1) \quad E\{\dots, x_{m-1}, x_m, x_n\} = x_m^6$$

whenever $m < n$. Then $E\{x_m\} = a$ is independent of m , and

$$(1.2.2) \quad \dots \leq E\{(x_m - a)^2\} \leq E\{(x_{m+1} - a)^2\} \leq \dots$$

If the $\{x_n\}$ are defined for all negative integers,

$$(1.2.3) \quad \lim_{m \rightarrow -\infty} x_m = x_{-\infty}$$

exists with probability 1 and

$$(1.2.4) \quad \lim_{m \rightarrow \infty} E\{(x_{-\infty} - x_m)^2\} = 0.$$

If the $\{x_n\}$ are defined for all positive integers, and if the dispersions in (1.2.2) form a bounded sequence,

$$(1.2.3') \quad \lim_{m \rightarrow \infty} x_m = x_{\infty}$$

exists with probability 1, and

$$(1.2.4') \quad \lim_{m \rightarrow \infty} E\{(x_{\infty} - x_m)^2\} = 0.^7$$

It follows from (1.2.1) that

$$(1.2.5) \quad E\{x_n\} = E\{E\{x_m; x_n\}\} = E\{x_m\}.$$

Hence $\dots = E\{x_0\} = E\{x_1\} = \dots$. It will be no restriction to assume from now on that

$$\dots = E\{x_0\} = E\{x_1\} = \dots = 0.$$

It also follows from (1.2.1) that

$$(1.2.6) \quad E\{x_m x_n\} = E\{E\{x_m; x_m x_n\}\} = E\{x_m E\{x_m; x_n\}\} = E\{x_m^2\}.^8$$

Using this equation,

$$(1.2.7) \quad E\{x_n^2\} = E\{[(x_n - x_m) + x_m]^2\} = E\{(x_n - x_m)^2\} + E\{x_m^2\},$$

and the dispersions of the x_n thus form a monotone non-decreasing sequence.

⁶ The conditional expectation of a chance variable y for given values of a chance variable η will be denoted by $E\{\eta; y\}$.

⁷ Much of this theorem remains true if (1.2.1) is true but only the first moments of the x_n are supposed finite, no other hypothesis being made on their distributions. Cf. Doob, *Am. Math. Soc. Trans.*, Vol. 47 (1940), pp. 458-460.

⁸ Alternatively, (1.2.1) implies that $x_n - x_m$ is uncorrelated with x_m . Then $E\{x_m x_n\} = E\{x_m^2 + (x_n - x_m)x_m\} = E\{x_m^2\}$.

Finally, using (1.2.6),

$$(1.2.8) \quad E\{(x_{m+1} - x_m)(x_{n+1} - x_n)\} = 0.$$

The series

$$(1.2.9) \quad \sum_m (x_{m+1} - x_m)$$

is therefore a series of mutually independent chance variables. According to a well known theorem of Kolmogoroff, a sequence of mutually independent chance variables converges with probability 1 if the means and dispersions form a convergent series. The present theorem follows at once from Kolmogoroff's theorem.

COROLLARY. Let x be a one-dimensional Gaussian chance variable and let

$$(1.2.10) \quad \begin{array}{c} \dots\dots\dots \\ x_{01}, x_{02}, \dots \\ x_{11}, x_{12}, \dots \\ x_{21}, x_{22}, \dots \\ \dots\dots\dots \end{array}$$

be sequences of one-dimensional Gaussian chance variables with the property that if $\nu \geq 1$, the multivariate distribution of $x, x_{\nu 1}, \dots, x_{\nu \nu}$ is Gaussian, and suppose that each variable $x_{m\nu}$ is a member of every later sequence. Then

$$(1.2.11) \quad \begin{aligned} \lim_{m \rightarrow -\infty} E\{x_{m1}, x_{m2}, \dots; x\} &= x_- \\ \lim_{m \rightarrow +\infty} E\{x_{m1}, x_{m2}, \dots; x\} &= x_+ \end{aligned}$$

exist with probability 1, and in the mean, and

$$(1.2.12) \quad x_+ = E\{x_{mn}, m = 0, \pm 1, \dots, n = 1, 2, \dots; x\}.$$

It will first be shown that the sequence $\{x_m\}$, where

$$(1.2.13) \quad x_m = E\{x_{m1}, x_{m2}, \dots; x\},$$

has the property demanded in the theorem. In fact, from the definition of conditional expectation, the difference $x - x_n$ has expectation zero and is independent of the variables $\{x_{mj}\}$ for $m \leq n$, and therefore of the variables \dots, x_{n-1}, x_n . Hence

$$(x - x_n) - (x - x_{n+1}) = x_{n+1} - x_n$$

has expectation zero and is independent of the variables \dots, x_{n-1}, x_n . Therefore the sequence $\{x_{n+1} - x_n\}$ is a sequence of mutually independent chance variables with vanishing expectations. This implies (1.2.1) if $m < n$ because

$$(1.2.14) \quad \begin{aligned} E\{\dots, x_m; x_n\} &= E\left\{\dots, x_m; x_m + \sum_{j=m}^{n-1} (x_{j+1} - x_j)\right\} = E\{\dots, x_m; x_m\} \\ &+ \sum_{j=m}^{n-1} E\{\dots, x_m; x_{j+1} - x_j\} = x_m. \end{aligned}$$

Let a be the common value of $E\{x\}$, $E\{x_m\}$. Since $x - x_m$ is independent of x_m ,

$$(1.2.15) \quad E\{x_m - a\}^2 + E\{(x - x_m)^2\} = E\{(x - a)^2\}.$$

Hence the sequence of dispersions of the x_m is bounded and according to Theorem 1.2 the limits x_- and x_+ in (1.2.1) exist with probability 1. Since $x - x_m$ has expectation zero and is independent of x_m , for $m \leq n$, $x - x_+$ also has expectation zero and is independent of x_m , for all m , that is (1.2.12) is true.

The simplest non-trivial special case of this theorem is the following: Let x_1, x_2, \dots, x be one-dimensional Gaussian chance variables with the property that if $\nu \geq 1$ the multivariate distribution of x, x_1, \dots, x_ν is Gaussian. Then

$$(1.2.16) \quad \lim_{n \rightarrow \infty} E\{x_1, \dots, x_n, x\} = E\{x_1, x_2, \dots; x\},$$

with probability 1, and this limit is also a limit in the mean.

As stated, the theorem and corollary are true without the hypothesis that the chance variables concerned are Gaussian. (The existence of second moments must be assumed if the limits are to exist as limits in the mean.) They are stated for Gaussian variables because the proof is simple in that case, and because that is sufficient for the purposes of this paper.

In discussing t.h.G. processes whose parameter t is not restricted to be integral, the usual continuity hypothesis will be made. It will be supposed that $R(t)$ is continuous at $t = 0$:

$$(1.3.1) \quad \lim_{t \rightarrow 0} [R(t) - R(0)] = -\frac{1}{2} \lim_{t \rightarrow 0} E\{[x(t) - x(0)]^2\} = 0.$$

It is then easily concluded that $R(t)$ is everywhere continuous.

In the continuous parameter case, it would be useful to have the conditions on $R(t)$ necessary and sufficient for the continuity in t of the chance variables $x(t)$ and for the existence of the derivative. No set of necessary and sufficient conditions for the continuity of $x(t)$ is known, although the fact of continuity will not be difficult to prove in the special cases to be considered in §4. The conditions for the existence of $x'(t)$ are quite simple, and will be given in Theorem 1.4.

The spectral function of a one-dimensional t.h.G. process will play an essential role in some of the theorems to be discussed below. If $R(n)$ is the correlation function of a one-dimensional t.h.G. process, $R(n)$ can be expressed in either of the forms

$$(1.3.2) \quad R(n) = \int_0^\pi \cos n\lambda \, dF(\lambda) \quad n = 0, \pm 1, \dots,$$

$$(1.3.2') \quad R(n) = \int_{-\pi}^\pi e^{in\lambda} \, dG(\lambda) \quad n = 0, \pm 1, \dots,$$

where, $F(\lambda)$, called the spectral function of the process, and $G(\lambda)$, called the complex spectral function of the process, are real monotone non-decreasing functions satisfying the following conditions:

$$\begin{aligned}
 (1.3.3) \quad & F(0) = 0 & G(-\pi) = 0 \\
 & F(\lambda-) = F(\lambda), \quad 0 < \lambda < \pi, & G(\lambda-) = G(\lambda), \quad -\pi < \lambda < \pi \\
 & & G(\lambda) - G(0+) = G(0) - G(-\lambda+) \\
 (1.3.4) \quad & F(\pi) = G(\pi) \\
 & F(\lambda) = G(\lambda) - G(-\lambda+) = 2G(\lambda) - G(0) - G(0+), \quad 0 < \lambda < \pi \\
 & F'(\lambda) = 2G'(\lambda)^*
 \end{aligned}$$

The last equation of course holds only at points where the derivatives exist. The forms (1.3.2), (1.3.2') are derived trivially from each other. The correlation function determines the spectral functions uniquely, if the latter are supposed to satisfy (1.3.3). In fact, at the points of continuity of $F(\lambda)$, $G(\lambda)$:

$$\begin{aligned}
 (1.3.5) \quad & F(\lambda) = \frac{\lambda R(0)}{\pi} + \frac{2}{\pi} \sum_1^{\infty} R(n) \frac{\sin \frac{n\lambda}{2}}{n} \\
 & G(\lambda) = \frac{(\lambda + \pi)R(0)}{2\pi} + \frac{1}{2\pi i} \lim_{r \rightarrow \infty} \sum_{\substack{n=1 \\ n \neq 0}}^r R(n) \frac{e^{in\lambda}}{n}.
 \end{aligned}$$

Conversely if any $F(\lambda)$ or $G(\lambda)$ satisfying the stated conditions is used to determine an $R(n)$ by means of (1.3.2) or (1.3.2'), $R(n)$ is the correlation function of a t.h.G. process. The representation of $R(n)$ in terms of $G(\lambda)$ is frequently more convenient than that in terms of $F(\lambda)$, because of the simple properties of the exponential function. The following relation, which will be used below, exhibits the elegance attained by the use of $G(\lambda)$:

$$\begin{aligned}
 (1.3.6) \quad & E\left[\left[\sum_m c'_m x(m)\right] \cdot \left[\sum_n c''_n x(n)\right]\right] = \sum_{m,n} c'_m c''_n R(m-n) \\
 & = \int_{-\pi}^{\pi} \left(\sum_m c'_m e^{im\lambda}\right) \left(\overline{\sum_n c''_n e^{in\lambda}}\right) dG(\lambda).
 \end{aligned}$$

The correlation function of a one-dimensional continuous parameter t.h.G. process can be represented in either of the following forms:

$$(1.3.7) \quad R(t) = \int_0^{\infty} \cos t\lambda \, dF(\lambda)$$

$$(1.3.7') \quad R(t) = \int_{-\infty}^{\infty} e^{it\lambda} \, dG(\lambda)$$

* H. Wold, *A Study in the Analysis of Stationary Time Series*, Uppsala, (1938), p. 66.

where the spectral function $F(\lambda)$ and the complex spectral function $G(\lambda)$ are monotone non-decreasing and satisfy the conditions

$$\begin{aligned}
 (1.3.8) \quad & F(0) = 0 & G(-\infty) &= 0 \\
 & F(\lambda-) = F(\lambda), \quad 0 < \lambda < \infty & G(\lambda-) &= G(\lambda) \\
 & & G(\lambda) - G(0+) &= G(0) - G(-\lambda+) \\
 (1.3.9) \quad & F(\infty) = G(\infty) \\
 & F(\lambda) = G(\lambda) - G(-\lambda+) = 2G(\lambda) - G(0) - G(0+), \quad 0 < \lambda < \infty, \\
 & F'(\lambda) = 2G'(\lambda).
 \end{aligned}$$

The last equation of course only holds at points where the derivatives exist. The correlation function $R(t)$ determines the spectral functions uniquely if the latter are supposed to satisfy (1.3.8). In fact, at the points of continuity of $F(\lambda)$, $G(\lambda)$:

$$\begin{aligned}
 (1.3.10) \quad & F(\lambda) = \frac{2}{\pi} \int_0^\infty R(t) \frac{\sin t\lambda}{t} dt \\
 & G(\lambda) = \lim_{\tau \rightarrow \infty} \frac{1}{2\pi i} \int_{-\tau}^\tau R(t) \frac{e^{i\lambda} - 1}{t} dt + G(0).
 \end{aligned}$$

THEOREM 1.4 Let $\{x(t)\}$ be the variables of a one-dimensional continuous parameter t.h.G. process with correlation function $R(t)$ and spectral function $F(\lambda)$. If

$$(1.4.1) \quad \int_0^\infty \lambda^2 dF(\lambda) < \infty$$

then

- (i) $R'(t)$, $R''(t)$ exist and are continuous, and $R'(0) = 0$;
- (ii) $x(t)$ is an absolutely continuous function of t , with probability 1;
- (iii) for each t ,

$$(1.4.2) \quad \lim_{h \rightarrow 0} \frac{x(t+h) - x(t)}{h} = x'(t)$$

exists, with probability 1, and this convergence is also true in the mean:

$$(1.4.3) \quad \lim_{h \rightarrow 0} E \left\{ \left[\frac{x(t+h) - x(t)}{h} - x'(t) \right]^2 \right\} = 0;$$

(iv) the $x'(t)$ process is a t.h.G. process, with correlation function $-R''(t)$ and spectral function $\int_0^\lambda \lambda^2 dF(\lambda)$.

Conversely if

$$(1.4.4) \quad \liminf_{h \rightarrow 0} \frac{R(0) - R(h)}{h^2} = \frac{1}{2} \liminf_{h \rightarrow 0} E \left\{ \left[\frac{x(h) - x(0)}{h} \right]^2 \right\} < \infty,$$

then (1.4.1) is true.

This theorem is due to Slutsky. The proof will be sketched here, for completeness. (The hypothesis that the process is a Gaussian process is immaterial, since only the second moments are involved in the proof.)

Proof of (i). If the integral (1.4.1) exists, $R'(t)$, $R''(t)$ can be obtained by differentiating under the integral sign in (1.2.2):

$$(1.4.5) \quad \begin{aligned} R'(t) &= - \int_0^\infty \lambda \sin \lambda t \, dF(\lambda) \\ R''(t) &= - \int_0^\infty \lambda^2 \cos \lambda t \, dF(\lambda). \end{aligned}$$

Then $R'(t)$, $R''(t)$ are continuous functions, and $R'(0) = 0$.

Proof of (ii), (iii), (iv). The quantity

$$(1.4.6) \quad E \left\{ \left[\frac{x(t+h_1) - x(t)}{h_1} - \frac{x(t+h_2) - x(t)}{h_2} \right]^2 \right\}$$

can be evaluated in terms of the correlation function $R(t)$, and approaches 0 with h_1, h_2 , if the second derivative $R''(t)$ exists. There is therefore a chance variable $y(t)$ to which the difference quotient converges in the mean:

$$(1.4.7) \quad \lim_{h \rightarrow 0} E \left\{ \left[\frac{x(t+h) - x(t)}{h} - y(t) \right]^2 \right\} = 0.$$

The $y(t)$ process is a t.h.G. process. Moreover the equation

$$(1.4.8) \quad E\{x(s)x(s+t)\} = R(t)$$

can be differentiated to give

$$(1.4.9) \quad E\{x(s)y(s+t)\} = E\{x(s-t)y(s)\} = R'(t)$$

and this in turn when differentiated becomes

$$(1.4.10) \quad E\{y(s-t)y(s)\} = E\{y(s)y(s+t)\} = -R''(t).$$

Hence the $y(t)$ process has correlation function $-R''(t)$. Finally,

$$(1.4.11) \quad \begin{aligned} E \left\{ \left[x(t) - x(0) - \int_0^t y(s) \, ds \right]^2 \right\} &= E\{[x(t) - x(0)]^2\} \\ &+ \int_0^t \int_0^t E\{y(s)y(s')\} \, ds \, ds' - 2 \int_0^t E\{[x(t) - x(0)]y(s)\} \, ds = 0, \end{aligned}$$

(evaluating the right side of (1.4.11) in terms of $R(t)$, $R'(t)$, $R''(t)$). Thus $x(t)$ is absolutely continuous, with probability 1, and $y(t)$ is the derived function $x'(t)$. Hence $x'(t)$ exists for almost all t , with probability 1.¹⁰ It follows (Fubini's theorem) that the limit in (1.4.2) exists for each t , with probability 1, except possibly for a λ -set of Lebesgue measure 0. Since the process is t.h., the

¹⁰ For the exact meaning and measure-theoretic justification for statements of this type, see Doob, *Am. Math. Soc. Trans.*, Vol. 42 (1937), pp. 107-40

exceptional set must be either empty or the whole t -line. The exceptional set is therefore empty.

Conversely if (1.4.4) is true, (1.4.1) follows at once from (1.3.7).

It will be convenient to use condensed notation below. If $x: (x_1, \dots, x_N)$, $y: (y_1, \dots, y_N)$ are N -dimensional vectors and if $A: (a_{ij})$ is an N -dimensional square matrix, $x \cdot y$ will denote the matrix $(x_i y_j)$, Ax the vector with components $\sum_j a_{ij} x_j$; and (x, y) the number $\sum_i x_i y_i$. The adjoint matrix (a_{ij}^*) : $a_{ij}^* = a_{ji}$, will be denoted by A^* . Throughout this paper, the chance variables will be real-valued, but it will be convenient to use complex constant vectors. The identity matrix will be denoted by I . It will be convenient to denote the i, j th term of the matrix A by $(A)_{ij}$. The following equations will be used frequently:

$$Ax \cdot By = A(x \cdot y)B^*, \quad (Ax, y) = (x, A^*y).$$

If x is a chance variable, it is clear that $E\{x \cdot x\}$ is a symmetric non-negative definite matrix.

The simplest Gaussian processes are those in which the distribution of future states is based not on the complete past, but only on the immediate present. The precise definition of this (Markoff) property is the following.

(C) If $t_1 < \dots < t_{\nu+1}$ the conditional distribution of $x(t_{\nu+1})$ for given values of $x(t_1), \dots, x(t_\nu)$ depends only on the value assigned to $x(t_\nu)$. The conditional distribution of $x(t_{\nu+1})$ for given values of $x(t_1), \dots, x(t_\nu)$ will then be simply the conditional distribution of $x(t_{\nu+1})$ for the assigned value of $x(t_\nu)$.

The processes to be discussed in this paper are the processes with properties (A), (B), (C): temporally homogeneous Gaussian Markoff (t.h.G.M.) processes. The properties of t.h.G.M. processes will also be used to derive properties of the most important simple types of one-dimensional t.h.G. processes—those with rational spectral density functions. Some of the results are contained implicitly in the work of previous writers, but the presentation of the results has in all cases been chosen to stress their specific probability significance, and may therefore appeal even to readers familiar with previous work.

The condition (C) on a Gaussian process is equivalent to the condition (C') that if $t_1 < \dots < t_{\nu+1}$

$$(1.5.1) \quad E\{x(t_1), \dots, x(t_\nu); x(t_{\nu+1})\} = E\{x(t_\nu); x(t_{\nu+1})\}.$$

In fact (C) is at least as strong as (C'). Conversely if (C') is true,

$$(1.5.2) \quad \begin{aligned} x(t_{\nu+1}) &= x(t_{\nu+1}) - E\{x(t_\nu); x(t_{\nu+1})\} + E\{x(t_\nu); x(t_{\nu+1})\} \\ &= y + E\{x(t_\nu); x(t_{\nu+1})\}, \end{aligned}$$

where y is a Gaussian chance variable with mean 0 uncorrelated with and therefore independent of $x(t_1), \dots, x(t_\nu)$, and the last term of (1.5.2) is simply a multiple of $x(t_\nu)$. Then the conditional distribution of $x(t_{\nu+1})$ for given $x(t_1), \dots, x(t_\nu)$ is a Gaussian variable, with mean $E\{x(t_\nu); x(t_{\nu+1})\}$ and dispersion that of y . Since this conditional distribution depends only on $x(t_\nu)$, property (C')

implies property (C). Hence these properties are equivalent. The condition (C') can be written in the form

$$(1.5.3) \quad E\{x(\tau), \tau \leq s; x(s+t)\} = E\{x(s); x(s+t)\}, \quad t > 0.$$

In many applications the stochastic processes either have this property already or will have it if the dimensionality of the processes is increased by the adjunction of auxiliary chance variables. In the latter case the process is called a component process of a t.h.G.M. process. Component processes are discussed in detail below. If a process is a t.h.G. process, the right side of (1.5.3) is a linear transformation (depending only on t) of $x(s)$:

$$(1.5.4) \quad E\{x(s); x(s+t)\} = A(t)x(s), \quad t > 0.$$

The matrix function $A(t)$ will be called the transition matrix function. It satisfies the equation (obtained by performing the operation $E\{x(s) \cdot \}$ on both sides of (1.5.4))

$$(1.5.5) \quad R(t) = R(0)A(t)^*, \quad t > 0,$$

but is otherwise unrestricted since if (1.5.5) is true, the difference $x(t) - A(t)x(s)$ is uncorrelated with and therefore independent of $x(s)$. In many applications the elements of $R(t)$ will vanish identically except in square matrices down the main diagonal. If this is true, $A(t)$ can also be assumed in this form.

If the variables $\{x(t)\}$ determine an N -dimensional t.h.G.M. process, and if B is a non-singular N -dimensional square matrix, the variables $\{Bx(t)\}$ also determine a t.h.G.M. process. Two processes connected in this way will be called equivalent. If two t.h.G. processes are equivalent, and if one is a Markoff process, the other must be also. If there is a change of variable

$$(1.5.6) \quad y(t) = Bx(t).$$

taking the t.h.G.M. $x(t)$ process with transition matrix $A(t)$ and correlation matrix $R(t)$ into the equivalent $y(t)$ process with transition matrix $A_1(t)$ and correlation matrix $R_1(t)$, then

$$(1.5.7) \quad A_1(t) = BA(t)B^{-1}, \quad R_1(t) = BR(t)B^*.$$

If $\{x(t)\}$, $\{y(t)\}$, $\{z(t)\}$ determine t.h.G. processes of dimensions α , β and $\alpha + \beta$ respectively, if the process determined by

$$\{x_1(t), \dots, x_\alpha(t), \quad y_1(t), \dots, y_\beta(t)\}$$

is equivalent to the $z(t)$ process, and if every $x(s)$ is independent of every $y(t)$, the $z(t)$ process will be called the direct product of the $x(t)$ and $y(t)$ processes. The extension of the definition to direct products of more than two processes is clear. If the $x(t)$ and $y(t)$ processes are Markoff processes, their direct product is also a Markoff process. Conversely if the $z(t)$ process is a Markoff process, the factor processes must also be Markoff processes. The following facts about matrices will be used below. If A is any N -dimensional matrix, there is a non-

singular N -dimensional matrix B such that $B^{-1}AB$ is in Jordan canonical form: the elements of $B^{-1}AB$ vanish except for those in certain submatrices down the main diagonal. Each of these submatrices has the form

$$(1.5.8) \quad \begin{pmatrix} \lambda & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & \lambda & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & \lambda \end{pmatrix}$$

or simply (λ) if it is one-dimensional. The λ 's are the characteristic values of A , that is the roots of the characteristic equation $\det. |A - \lambda I| = 0$, and the sum of the dimensions of the submatrices with a given λ is the multiplicity of λ as a root of this equation. The matrix A is said to have simple elementary divisors corresponding to a given root λ of the characteristic equation if the submatrices in (1.5.7) with that λ are all of dimension 1. Thus orthogonal matrices, symmetric matrices, and skew symmetric matrices have only simple elementary divisors, since they can be put in diagonal form, (with λ 's of modulus 1, real, pure imaginary, respectively). The transformation B and the λ 's may not be real. If A is real, however, there is a real matrix B such that the elements of $B^{-1}AB$ vanish except for square submatrices down the main diagonal, and the characteristic roots of different submatrices are neither equal nor conjugate imaginary.

The powers of a matrix in Jordan canonical form are easily calculated using the fact that

$$(1.5.9) \quad \begin{pmatrix} \lambda & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & \lambda & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & \lambda \end{pmatrix}^n = \begin{pmatrix} \lambda^n & 0 & \cdot & \cdot & \cdot & 0 \\ n\lambda^{n-1} & \lambda^n & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & n\lambda^{n-1} & \lambda^n \end{pmatrix}.$$

It follows that in the general case the elements $(A^n)_{ij}$ are linear combinations of λ_j^n , $n\lambda_j^{n-1}$, etc., where $\lambda_1, \lambda_2, \dots$ are the characteristic values of A . Hence if $(A^n)_{ij} \rightarrow 0$ as $n \rightarrow \infty$, the approach must be exponential. The terms of A^n certainly go to 0 if all the characteristic values of A have modulus less than 1.

The matrix e^A is defined by the usual series formula for the exponential function. If A has the form (1.5.8), e^{tA} can be calculated using (1.5.9):

$$(1.5.10) \quad e^{tA} = \begin{pmatrix} e^{t\lambda} & 0 & \cdot & \cdot & \cdot & 0 \\ te^{t\lambda} & e^{t\lambda} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & te^{t\lambda} & e^{t\lambda} \end{pmatrix}.$$

It follows that in the general case the elements $(e^{tA})_{ij}$ are linear combinations of $e^{t\lambda_j}$, $te^{t\lambda_j}$, etc., where $\lambda_1, \lambda_2, \dots$ are the characteristic values of A . If $(e^{tA})_{ij} \rightarrow 0$ as $t \rightarrow \infty$ the approach must be exponential. The terms of e^{tA} certainly go to 0 if all the characteristic values of A have negative real parts.

A t.h.G.M. process will be called deterministic if the least squares prediction of $x(s+t)$ for given $x(s)$, ($t > 0$), that is $E\{x(s+t); x(s)\}$ is always correct:

$$(1.6.1) \quad x(s+t) = A(t)x(s) \quad t > 0,$$

with probability 1.

The following classification of deterministic processes will be useful later. It will be shown that any t.h.G.M. process is the direct product of processes of the following four deterministic types, and of a factor process with no deterministic factors.

$M(0)$. Let $\{x(t)\}$ be the variables of a one-dimensional t.h.G.M. process, with $x(t) = 0$ with probability 1. (The chance variable which is 0 with probability 1 is considered as a Gaussian variable with mean 0 and dispersion 0.) The correlation function of the process vanishes identically.

$M(1)$. A one-dimensional t.h.G.M. process which satisfies

$$(1.6.2) \quad x(t) = x(0), \quad E\{x(t)\} = 0, \quad E\{x(t)^2\} > 0,$$

will be called a process of type $M(1)$. The correlation function $R(t)$ is positive and independent of t .

$M(-1)$. A one-dimensional t.h.G.M. process with an integral-valued parameter n , satisfying

$$(1.6.3) \quad \dots = x(-1) = -x(0) = x(1) = \dots \quad E\{x(n)\} = 0, \quad E\{x(n)^2\} > 0$$

will be called a process of type $M(-1)$. The correlation function is alternately positive and negative: $R(n) = (-1)^n R(0)$.

$M(e^{i\theta})$. A two-dimensional t.h.G.M. process with

$$(1.6.4) \quad \begin{aligned} E\{x_1(0)\} &= 0, & E\{x_1(0)^2\} &= \sigma^2 > 0, & E\{x_1(0)x_2(0)\} &= 0, \\ x_1(t) &= x_1(0) \cos t\theta - x_2(0) \sin t\theta \\ x_2(t) &= x_1(0) \sin t\theta + x_2(0) \cos t\theta, \end{aligned}$$

will be called a process of type $M(e^{i\theta})$. The correlation function is given by

$$(1.6.5) \quad R(t) = \begin{pmatrix} \sigma^2 \cos t\theta & \sigma^2 \sin t\theta \\ -\sigma^2 \sin t\theta & \sigma^2 \cos t\theta \end{pmatrix}.$$

A process with variables $\{x(t)\}$ will be called degenerate if there are constants c_1, \dots, c_N not all 0, such that

$$(1.7.1) \quad \sum_j c_j x_j(t) = 0$$

with probability 1, for all t . Equation (1.7.1) is true if and only if

$$(1.7.2) \quad E\{[\sum_j c_j x_j(t)]^2\} = \sum_{j,k} (R(0))_{jk} c_j c_k = 0$$

that is if and only if the correlation matrix $R(0)$ is singular. If a non-degenerate process is a direct product of factor processes, the latter are also non-degenerate. The only degenerate one-dimensional process is that of type $M(0)$.

2. The structure of degenerate and deterministic processes

THEOREM 2.1. *Every degenerate t.h.G.M. process is the direct product of processes of type $M(0)$ and (in some cases) of a non-degenerate t.h.G.M. process.*

In proving this theorem, it can be supposed that the original process has been replaced by an equivalent process, if necessary, so that the symmetric non-negative definite matrix $R(0)$ is in diagonal form, with only 0's and 1's down the main diagonal, say 0 to the ν th place and 1 thereafter. Then $x_j(t) = 0$, when $j \leq \nu$ and the process is obviously the direct product of ν processes of type $M(0)$ and an $(N - \nu)$ -dimensional non-degenerate process.

THEOREM 2.2. *Let $\{x(t)\}$ be the variables of a deterministic t.h.G.M. process, with correlation function $R(t)$.*

(i) *The process is the direct product of factor processes of types $M(0)$, $M(\pm 1)$, $M(e^{i\theta})$*

(ii) *If the parameter t of the process is restricted to the integers, there is a non-singular matrix A such that*

$$(2.2.1) \quad x(n) = A^n x(0),$$

$$(2.2.2) \quad R(n) = R(0)A^{*n}, \quad n = 0, \pm 1, \pm 2, \dots$$

$$(2.2.3) \quad R(0) = AR(0)A^*.$$

The transition matrix A is the transform BOB^{-1} of an orthogonal matrix O . If the process is non-degenerate, A is uniquely determined.

(ii') *If the parameter of the process runs through all real numbers, there is a matrix Q such that*

$$(2.2.1') \quad x(t) = e^{iQ} x(0), \quad -\infty < t < \infty,$$

$$(2.2.2') \quad R(t) = R(0)e^{iQ^*t},$$

$$(2.2.3') \quad QR(0) + R(0)Q^* = 0.$$

The matrix Q is the transform BKB^{-1} of a skew symmetric matrix K . If the process is non-degenerate, Q is uniquely determined

(iii) *Conversely if $R(0)$ is any symmetric non-negative definite matrix, and if $A(Q)$ is any matrix satisfying (2.2.3) ((2.2.3')), where A is non-singular, there is a deterministic t.h.G.M. process, with correlation function given by (2.2.2) ((2.2.2')) and satisfying (2.2.1) ((2.2.1')).*

In proving (i) (ii) and (ii') it will be permissible to go to processes equivalent to the original one, if convenient. Moreover if the given process can be expressed as a direct product, it will be sufficient to prove (i) (ii) and (ii') for each factor. Since (i) (ii) and (ii') are certainly true for processes of type $M(0)$ (with A in (ii) the identity, and Q in (ii') the null matrix), and since according to Theorem 2.1, processes of type $M(0)$ can be factored out of the given process to leave a non-degenerate remaining factor, if any, it will be sufficient to prove (i) (ii) and (ii') for non-degenerate processes.

Proof (t integral) of (i) and (ii) for non-degenerate processes. If the process determined by $\{x(n)\}$ is deterministic, (1.6.1) is true. Hence

$$(2.2.4) \quad x(\nu + 1) = Ax(\nu).$$

Then (2.2.1) is true for $n \geq 0$, and will be established for all n as soon as it is shown that A is non-singular. Using (2.2.1),

$$(2.2.5) \quad R(n) = E\{x(0) \cdot x(n)\} = E\{x(0) \cdot A^n x(0)\} = R(0)A^{*n}, \quad n \geq 0,$$

and

$$(2.2.6) \quad R(0) = E\{x(1) \cdot x(1)\} = E\{Ax(0) \cdot Ax(0)\} = AR(0)A^*.$$

Under the present hypotheses, $R(0)$ is non-singular. Then A is determined uniquely by (2.2.5) with $n = 1$, and A cannot be singular because of (2.2.6). There is an equivalent process in which $R(0)$ is the identity. Considering this process, (2.2.6) becomes $I = AA^*$, so that A is orthogonal. Finally there is an equivalent process (obtained by an orthogonal change of variables) in which $R(0)$ is still the identity and the matrix A is in the (real) normal form of orthogonal matrices: all the elements of A are 0 except for two-dimensional rotation matrices or 1's or -1's down the main diagonal. It is now obvious that the process is the direct product of processes of types $M(\pm 1)$, $M(e^{i\theta})$.

Proof (t continuously varying) of (i) and (ii') for non-degenerate processes. If the t.h.G.M. process determined by $\{x(t)\}$ is deterministic, (1.6.1) is true. Hence

$$(2.2.5') \quad R(t) = E\{x(s) \cdot x(s+t)\} = R(0)A(t)^*$$

$$(2.2.6') \quad R(0) = E\{x(s+t) \cdot x(s+t)\} = A(t)R(0)A(t)^*.$$

The matrix $A(t)$ is uniquely determined by (2.2.5') since $R(0)$ is non-singular. It then follows from (1.6.1) that

$$(2.2.7) \quad A(s+t) = A(s)A(t).$$

The continuity hypothesis (1.3.1) becomes

$$(2.2.8) \quad \lim_{t \rightarrow 0} R(0)A(t)^* = R(0),$$

which implies that

$$(2.2.9) \quad \lim_{t \rightarrow 0} A(t) = I.$$

It is well known that any solution of (2.2.7) and (2.2.9) can be written in the form $A(t) = e^{tQ}$. If now the right side of (2.2.6') is expanded in powers of t and the coefficient of t is set equal to 0, the resulting equation is (2.2.3'). It can be supposed, going to an equivalent process if necessary, that $R(0)$ is the identity. Then $A(t)A(t)^* = I$, $Q + Q^* = 0$. An equivalent process can be chosen so

that $R(0)$ is still the identity, and so that Q is in the real canonical form of skew symmetric matrices: its elements vanish except for possible two rowed matrices

$$\begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix}$$

down the main diagonal. It is now clear that the non-degenerate process is a direct product of factors of type $M(e^{\theta})$ corresponding to two rowed matrices just described, and factors of type $M(1)$.

Proof of (iii). If $R(0)$ and $A(Q)$ satisfy the conditions of Theorem 1.2 (iii), choose $x(0)$ as any Gaussian variable with correlation matrix $R(0)$. Then define $x(n)$ by (2.2.1) ((2.2.1')). The resulting stochastic process is temporally homogeneous if and only if $E\{x(s) \cdot x(s+t)\}$ depends only on t . The details of the calculation will be carried out for only for t integral.

In the first place

$$(2.2.10) \quad E\{x(n) \cdot x(n+\nu)\} = E\{A^n x(0) \cdot A^{n+\nu} x(0)\} = A^n R(0) A^{n+\nu}.$$

Now (2.2.3) can be developed further:

$$(2.2.11) \quad R(0) = AR(0)A^* = A^2 R(0) A^{*2} = \dots$$

so that (2.2.10) reduces to

$$(2.2.12) \quad E\{x(n) \cdot x(n+\nu)\} = R(0) A^{n\nu}.$$

The process is thus temporarily homogeneous, and obviously satisfies the other parts of the definition of a deterministic t.h.G.M. process. Theorem 2.2 is now completely proved.

The restriction imposed on $R(0)$, $A(Q)$ by (2.2.3) ((2.2.3')) is quite loose. Given $R(0)$, there is always an $A(Q)$ satisfying (2.2.3) ((2.2.3')) for example the identity (null matrix). Given an A which is the transform of an orthogonal matrix (a Q which is the transform of a skew symmetric matrix) there is always a corresponding $R(0)$: In fact $A(Q)$ can be assumed to be orthogonal (skew symmetric) and the $R(0)$ can be taken as the identity.

3. T.H.G.M. processes with an integral valued parameter

In this section, the parameter t will range through all the integers. The condition (1.5.3) that a t.h.G. process be a t.h.G.M. process can be simplified in the integral parameter case. In fact it will be shown that it is sufficient if

$$(3.1.1) \quad E\{\dots, x(n-1), x(n); x(n+1)\} = E\{x(n); x(n+1)\},$$

$n = 0, \pm 1, \dots$

with probability 1. If (3.1.1) is true,

$$(3.1.2) \quad x(\nu+1) = Ax(\nu) + \eta(\nu)$$

where A is the transition matrix of the process and $\eta(\nu)$ has mean 0 and is independent of $\dots, x(\nu - 1), x(\nu)$. It follows that

$$(3.1.3) \quad x(n) = A^{n-m}x(m) + A^{n-m-1}\eta(m) + A^{n-m-2}\eta(m+1) + \dots + \eta(n-1).$$

The terms involving the $\eta(j)$ are all independent of $\dots, x(m-1), x(m)$. This equation therefore implies that

$$(3.1.4) \quad E\{\dots, x(m-1), x(m); x(n)\} = A^{n-m}x(m) = E\{x(m); x(n)\},$$

and (3.1.4) is precisely the condition that the process has the Markoff property.

The following lemma will be useful.

LEMMA 3.2. *Let $x: (x_1, \dots, x_N)$ be any Gaussian chance variable, with $E\{x\} = 0$. Then there is a uniquely determined symmetric non-negative definite matrix S , and a Gaussian chance variable y , such that*

$$(3.2.1) \quad E\{y \cdot y\} = I$$

and

$$(3.2.2) \quad x = Sy, \quad S^2 = E\{x \cdot x\}.$$

If $x = Sy$, and if S is symmetric, then the second equation in (3.2.2) is certainly true. It is easily seen, by examination of the characteristic values and vectors of the matrix $E\{x \cdot x\}$ that this matrix has a unique symmetric non-negative definite square root S . Hence if there is an S satisfying (3.2.2), there can be only one. The chance variables x_1, \dots, x_N can be written as linear combinations of N uncorrelated Gaussian chance variables ξ_1, \dots, ξ_N satisfying $E\{\xi \cdot \xi\} = I$:

$$(3.2.3) \quad x = A\xi.$$

If A is written in the polar form $A = SU$, where S is symmetric and non-negative definite and U is orthogonal, (3.2.3) becomes

$$(3.2.4) \quad x = SU\xi = Sy$$

where $y = U\xi$ satisfies (3.2.1).

It will be shown below that every t.h.G.M. process can be represented as the direct product of factors of certain types. The deterministic types have already been catalogued: $M(0)$, $M(\pm 1)$, $M(e^{i\theta})$. The non-deterministic factor type (integral valued parameter) will now be described.

M. Let $\{\eta(n)\}$ be a sequence of mutually independent N -dimensional Gaussian chance variables with 0 means and a common distribution function. Let A be any N -dimensional square matrix. Define $x(n)$ by

$$(3.3.1) \quad x(n) = \sum_{m=0}^{\infty} A^m \eta(n-m)$$

where it is supposed that¹¹ A is so chosen that the series converges with probability 1. This will be true, for example, if all the characteristic values of A have modulus less than 1, so that the terms of A^m go exponentially to 0¹². It will be shown below that it is no restriction to assume that A has this character. The variables $\{x(n)\}$ determine a t.h.G.M. process. Since $x(n) - Ax(n-1)$ is independent of $\dots, x(n-2), x(n-1)$, the $x(n)$ process is a Markoff process with transition matrix A :

$$(3.3.2) \quad E\{\dots, x(n-1); x(n)\} = Ax(n-1).$$

A process defined in this way will be called a process of type M . A non-singular change of variables $y(n) = Bx(n)$ leads to a process of the same type:

$$(3.3.3) \quad y(n) = \sum_{m=0}^{\infty} (BAB^{-1})^m B\eta(n-m).$$

It will sometimes be convenient to write a process of type M in a form slightly different from (3.3.1). Using Lemma 3.2 it is evident that there are Gaussian variables $\{\xi(n)\}$ satisfying

$$(3.3.4) \quad E\{\xi(n)\} = 0, \quad E\{\xi(m)\xi(n)\} = \delta_{m,n}I, \quad m, n = 0, \pm 1, \dots,$$

and a symmetric non-negative definite matrix S such that $\eta(n) = S\xi(n)$. Then $S^2 = E\{\xi(n)\xi(n)\}$ and

$$(3.3.5) \quad x(n) = \sum_{m=0}^{\infty} A^m S\xi(n-m).$$

Under the change of variable $y(n) = Bx(n)$, A becomes BAB^{-1} and S^2 becomes BS^2B^* .

The only condition on A required for convergence in (3.3.5) is that $A^m S \rightarrow 0$. It will now be shown that A can always be assumed to have only characteristic values of modulus less than 1, in the sense that there is an A with this property, and satisfying the equations

$$(3.3.6) \quad A^m S = \tilde{A}^m S, \quad m = 1, 2, \dots$$

It is no restriction, going to an equivalent process if necessary, to assume that the elements of A vanish except for those in two square submatrices down the main diagonal, where one submatrix A_1 has only characteristic values of modulus less than 1 and the other, A_2 , of modulus greater than or equal to 1. If the matrix S is written in terms of the corresponding submatrices:

$$(3.3.7) \quad A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \quad S = \begin{pmatrix} S_1 & S_3 \\ S_4 & S_2 \end{pmatrix}$$

¹¹ Throughout this paper, if A is any matrix, A^* is defined as the identity matrix I .

¹² We shall use repeatedly Kolmogoroff's theorem that an infinite series of mutually independent chance variables with zero means converges with probability 1 if the series of their dispersion is convergent. (Kolmogoroff only states the theorem in one dimension, but the extension to n dimension is trivial.) If a series of mutually independent Gaussian variables converges, the series of dispersions converges to the dispersion of the sum.

the condition on A implies that $A_1^m S_2 \rightarrow 0$. If it is shown that $S_2 = 0$, it will follow that $S_3 = S_4 = 0$, because S is symmetric and non-negative definite. The matrix \bar{A} will then be defined by

$$(3.3.8) \quad \bar{A} = \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix}$$

and A will satisfy (3.3.6) and will have only characteristic values of modulus less than 1. The problem has thus been reduced to the case where A_1 is absent: A only has characteristic values of modulus at least 1, and it must be proved that $A^m S \rightarrow 0$ implies that $S = 0$. The proof of this is immediate when A is put into its Jordan canonical form.

The symmetric non-negative definite matrix S satisfies the equations

$$(3.3.9) \quad R(0) = \sum_0 A^m S^2 A^{*m} = S^2 + AR(0)A^*.$$

THEOREM 3.4 *A direct product of processes of type M is also of type M . Conversely any factor process of a process of type M is itself of type M .*

The direct part of the theorem is obvious. To prove the converse, suppose an N -dimensional process of type M has an l -dimensional factor, corresponding to the variables $x_1(n), \dots, x_l(n)$. It can be supposed, that all factors of type $M(0)$ are separated out, so that there are indices $j, k: 1 \leq j \leq l \leq k < N$ such that the $\{x_1(n), \dots, x_j(n)\}$ and $\{x_{k+1}(n), \dots, x_N(n)\}$ processes are non-degenerate and that the variables $x_{j+1}(n), \dots, x_k(n)$ vanish identically. Making a change of variables, if necessary, it can be supposed that $R(0)$ has the form

$$(3.4.1) \quad R(0): \begin{pmatrix} j & k-j & N-k \\ I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I \end{pmatrix}$$

and that $R(n)$ has the blocks of zeros indicated in (3.4.1). Since $R(1) = R(0)A^*$, A must have the form

$$(3.4.2) \quad A: \begin{pmatrix} j & k-j & N-k \\ - & - & 0 \\ 0 & - & 0 \\ 0 & - & - \end{pmatrix}.$$

Then A^m will have this same form. Finally, because of (3.3.9), S^2 , and therefore S must have the form

$$(3.4.3) \quad S^2: \begin{pmatrix} - & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & - \end{pmatrix}.$$

Let A_0 be the matrix whose elements are the same as those of A except that the $(j+1)$ th to k th columns of A_0 vanish. Since

$$(3.4.4) \quad A_0^m S = A^m S, \quad m = 0, 1, \dots$$

it follows that

$$(3.4.5) \quad x(n) = \sum_{m=0}^{\infty} A_0^m S\xi(n-m).$$

It is now obvious that the $\{x_1(n), \dots, x_l(n)\}$ process is of type M .

THEOREM 3.5 *A non-degenerate process of type M has no deterministic factors.*

Any factor process is non-degenerate and of type M . To prove the theorem, it will therefore be sufficient to prove that the process itself cannot be deterministic. If it were, we should have

$$x(n) - Ax(n-1) = S\xi(n-1) = 0.$$

Then $S = 0$. But then the process is certainly degenerate, contrary to hypothesis.

THEOREM 3.6. (i) *Every t.h.G.M. process (discrete parameter) is the direct product of processes of type $M(0)$, $M(\pm 1)$, $M(e^{i\theta})$, M .*

(ii) *Let A be a transition matrix of a t.h.G.M. process, with variables $\{x(n)\}$. There are mutually independent Gaussian variables $\dots, \xi(0), \xi(1), \dots, \xi$ satisfying*

$$(3.6.1) \quad \begin{aligned} E\{\xi(n)\} &= E\{\xi\} = 0 \\ E\{\xi(n) \cdot \xi(n)\} &= E\{\xi \cdot \xi\} = I, \end{aligned}$$

and symmetric non-negative definite matrices S, T such that

$$(3.6.2) \quad x(n) = \sum_{m=0}^{\infty} A^m S\xi(n-m) + A^n T\xi, \quad n = 0, 1, \dots,$$

$$(3.6.3) \quad T^2 = AT^2A^*,$$

$$(3.6.4) \quad R(0) = \sum_{m=0}^{\infty} A^m S^2 A^{*m} + T^2 = AR(0)A^* + S^2,$$

where the series in (3.6.2) converges with probability 1. If A is non-singular, (3.6.2) holds for all n . The sum and last term in (3.6.2) are linear transformations of $x(n)$: (3.6.2) exhibits in part the decomposition into factor processes described in (i). The correlation function is given by

$$(3.6.5) \quad \begin{aligned} R(n) &= R(0)A^{*n} \\ R(-n) &= A^n R(0). \end{aligned} \quad n = 0, 1, \dots$$

(iii) *The transition matrix A is uniquely determined if and only if the process is non-degenerate. In any case, there is a transition matrix whose characteristic values are all of modulus less than or equal to 1, and whose characteristic values of modulus 1 correspond to simple elementary divisors. The transition matrix A furnishes the solution of the prediction problem of the process:*

$$(3.6.6) \quad E\{\dots, x(m-1), x(m); x(m+n)\} = A^n x(m), \quad n = 0, 1, \dots$$

The matrix S^2 , which is uniquely determined, measures the dispersion of $x(n+1)$ about its predicted value in terms of $x(n)$:

$$(3.6.7) \quad E\{[x(n+1) - Ax(n)]^2\} = S^2.$$

(iv) Conversely if A is a matrix with at least one characteristic value of modulus less than 1 or of modulus 1 and corresponding to a simple elementary divisor, A is the transition matrix of a t.h.G.M. process, with $R(0)$ not the null matrix. If all the characteristic values of A are as just described, A is the transition matrix of a non-degenerate t.h.G.M. process. If $R(0)$, S , A are matrices satisfying (3.6.4) with $R(0)$, S , symmetric and non-negative definite, there is a t.h.G.M. process whose variables can be written in the form (3.6.2) with the given $R(0)$, S , A .

This decomposition of a t.h.G.M. process into deterministic factors can be considered as a special case of the general decomposition theorem of Wold, which is applicable to all t.h.G. process.¹³ (Wold only considered the one-dimensional case.) The proof in the present special case is simpler, however, and illuminates the general case.

Proof of (i) and (ii). Equations (1.5.3) and (1.5.4), in the present case, lead to

$$(3.6.8) \quad E\{\dots, x(n-1), x(n); x(n+1)\} = E\{x(n); x(n+1)\} = Ax(n).$$

The first two terms are equal because the process has the Markoff property. The last term is linear in $x(n)$ because the process is Gaussian. The matrix A can be taken independent of n because the process is temporally homogeneous. Thus (3.6.8) involves the three fundamental properties of the $x(n)$ process. From the definition of conditional expectation, it follows that $x(n+1) - Ax(n)$ is independent of the chance variables $\dots, x(n-1), x(n)$. Hence the variables

$$\dots, [x(n) - Ax(n-1)], [x(n+1) - Ax(n)], \dots$$

are mutually independent. According to Lemma 3.2, there are mutually independent chance variables $\{\xi(n)\}$ satisfying

$$(3.6.9) \quad x(n) - Ax(n-1) = S\xi(n), \quad E\{\xi(n) \cdot \xi(n)\} = I, \quad E\{\xi(n)\} = 0,$$

where S is symmetric, non-negative definite, and satisfies (3.6.7). The matrix S^2 thus measures the dispersion of $x(n)$ about its predicted value $Ax(n-1)$.

The representation (3.6.2) can be obtained very simply. In fact

$$(3.6.10) \quad \begin{aligned} x(n) &= [x(n) - Ax(n-1)] + A[x(n-1) - Ax(n-2)] + \dots \\ &\quad + A^{n-\nu-1}[x(\nu+1) - Ax(\nu)] + A^{n-\nu}x(\nu) \\ &= \sum_{j=0}^{n-\nu-1} A^j S \xi(n-j) + A^{n-\nu}x(\nu), \end{aligned}$$

¹³ *A Study in the Analysis of Stationary Time Series*, Uppsala (1938), p. 89. See also Kolmogoroff, *Bull. Acad. Sci. URSS Ser. Math.*, Vol. 5 (1941), pp. 3-14 and *Bulletin Moskovskogo Gosudarstvennogo, Matematika*, Vol. 2 (1941), pp. 1-40, in whose papers the decomposition theorem is brought out in its full significance.

and it will be shown that when $\nu \rightarrow -\infty$ (3.6.10) leads to (3.6.2). Before going to the limit, however, we note that in (3.6.10) the sum is independent of the variables $\dots, x(\nu-1), x(\nu)$, so that

$$(3.6.11) \quad E\{\dots, x(\nu-1), x(\nu); x(n)\} = A^{n-\nu} x(\nu)$$

which is another way of writing (3.6.6). Moreover, using (3.6.11),

$$(3.6.12) \quad R(n-\nu) = E\{x(\nu) \cdot x(n)\} = R(0)A^{*n-\nu},$$

which is another way of writing (3.6.5). (The value of $R(n)$ for $n < 0$ is obtained using the fact that $R(-n) = R(n)^*$.) The last term in (3.6.10) is the conditional expectation of $x(n)$ for preassigned $\dots, x(\nu-1), x(\nu)$. It follows from the corollary to Theorem 1.2 that this conditional expectation converges with probability 1 when $\nu \rightarrow -\infty$, but this convergence will be proved directly in the present particular case.

From (3.6.10),

$$(3.6.13) \quad E\{x(n) \cdot x(n)\} = R(0) = \sum_{j=0}^{n-\nu-1} A^j S^2 A^{*j} + A^{n-\nu} R(0) A^{*n-\nu}.$$

The terms of the sum and the last term are all symmetric and non-negative definite matrices. It follows that there is convergence in (3.6.13) when $\nu \rightarrow -\infty$:

$$(3.6.14) \quad R(0) = \sum_{j=0}^{\infty} A^j S^2 A^{*j} + \lim_{m \rightarrow \infty} A^m R(0) A^{*m}.$$

The convergence of the series of dispersions in (3.6.14) implies that the series of chance variables in (3.6.2) converges, with probability 1. Then when $\nu \rightarrow -\infty$ (3.6.10) becomes

$$(3.6.15) \quad x(n) = \sum_{j=0}^{\infty} A^j S \xi(n-j) + z(n),$$

where

$$(3.6.16) \quad z(n) = \lim_{\nu \rightarrow -\infty} A^{n-\nu} x(\nu).$$

Since $x(n)$ is independent of $\xi(n+1), \xi(n+2), \dots$, $z(n)$ is independent of every $\xi(m)$. Moreover, writing $z(0) = T\xi$, where ξ satisfies (3.6.1) and T is symmetric and non-negative definite,

$$(3.6.17) \quad z(n) = A^n z(0) = A^n T \xi, \quad n \geq 0.$$

Thus (3.6.3) and (3.6.4) are satisfied. If A is non-singular, (3.6.17) will be correct for negative n also.

The decomposition of the process into factor processes of the types described in the theorem will be obtained by a detailed analysis of the significance of (3.6.2). Under the change of variable $y(n) = Bx(n)$, T^2 becomes BT^2B^* , and

A becomes BAB^{-1} . Making a suitable change of variables, if necessary, it can be supposed that A has the form

$$(3.6.18) \quad A: \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$$

where the characteristic values of A_1 have moduli unequal to 1 and those of A_2 have modulus 1. The matrix T^2 can be written in terms of submatrices of the same dimensions in a corresponding way:

$$(3.6.19) \quad T^2: \begin{pmatrix} T_1^2 & - \\ - & T_2^2 \end{pmatrix}$$

where T_1, T_2 are symmetric and non-negative definite. A further change of variables may be made, if necessary (transforming only the last n variables) preserving the forms (3.6.18) and (3.6.19) and transforming T_2 into the identity. Then using (3.6.3)

$$(3.6.20) \quad A_1 T_1^2 A_1^* = T_1^2, \quad A_2 A_2^* = I.$$

Hence A_2 is orthogonal. Developing (3.6.20) further, $A_1^m T_1^2 A_1^{*m} = T_1^2$, for all $m \geq 0$. When $m \rightarrow \infty$ in this equation, the terms in the matrix product on the left involve the m th power of the characteristic values of A_1 (all of modulus different from 1, by hypothesis). Then those characteristic values which actually appear can only be those of modulus less than 1, and the matrix on the left must go to 0 as $m \rightarrow \infty$; $T_1 = 0$. Since T is non-negative definite, T must have the simple form

$$(3.6.21) \quad T: \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}.$$

The matrix S can also be divided into corresponding submatrices:

$$(3.6.22) \quad S: \begin{pmatrix} S_1 & - \\ - & S_2 \end{pmatrix}.$$

The convergence of the series in (3.6.4) implies that

$$\lim_{m \rightarrow \infty} A_1^m S_2^2 A_1^{*m} = 0.$$

Since A_2 is orthogonal, this means that $S_2 = 0$, and since S is symmetric and non-negative definite, S has the form

$$(3.6.23) \quad S: \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix}.$$

It is now clear from (3.6.2) that the $x(n)$ process is the direct product of a process of type M and a deterministic process corresponding to the division of the variables determining the above submatrices. The deterministic factor process is the direct product of the elementary types already discussed. The variable $z(n)$ and the sum in (3.6.2) are linear transformations of $x(n)$.

Proof of (iii). If the process is non-degenerate, $R(0)$ is non-singular, and the transition matrix is determined uniquely by (3.6.5) with $n = 1$. If the process is degenerate, there will be one or more factor processes of type $M(0)$, and their transition matrices are quite unrestricted. In the non-degenerate case the characteristic values will be of modulus less than 1 (corresponding to a factor of type M , if one is present), or equal to 1 (corresponding to the factors of type $M(\pm 1)$, $M(e^{i\theta})$ making up the deterministic factor, if one is present), and in the latter case the elementary divisors are simple. If the process is degenerate, and if the part of A corresponding to the factors of type $M(0)$ is taken to be the identity, there will be simple elementary divisors corresponding to the characteristic value 1 for each such factor. The remaining statements of (iii) have already been proved.

Proof of (iv). Let A be a matrix with at least one characteristic value of modulus less than 1 or equal to 1 and corresponding to a simple elementary divisor. Then some transform BAB^{-1} has the form

$$(3.6.24) \quad \begin{pmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{pmatrix}$$

where A_1 (if present) has only characteristic values of modulus less than 1, A_2 (if present) is orthogonal, and both A_1, A_2 are not absent. For the purposes of the present proof it can be supposed that A is already in this form. Define S, T by

$$(3.6.25) \quad S: \begin{pmatrix} S_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad T: \begin{pmatrix} 0 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where the indicated submatrices of S and T are in the same positions as those of A , and where S_1 is any symmetric positive definite (and therefore non-singular) matrix of the proper dimension. The series in (3.6.4) converges and the first equation in (3.6.4) defines a matrix $R(0)$ which obviously satisfies the continued equality. If all the characteristic values of A are as described in the beginning of this paragraph, A_3 can be supposed absent. In this case

$$R(0) = S^2 + T^2 + \dots$$

is non-singular. The proof of the first two parts of (iv) has now been reduced to that of the last part. Suppose then that $R(0), A, S$ satisfy the hypotheses of the last part of (iv). Then

$$(3.6.26) \quad \begin{aligned} R(0) &= AR(0)A^* + S^2 \\ AR(0)A^* &= A^2R(0)A^{*2} + AS^2A^* \\ &\dots\dots\dots \\ A^{n-1}R(0)A^{*(n-1)} &= A^nR(0)A^{*n} + A^{n-1}S^2A^{*(n-1)}. \end{aligned}$$

Adding these equations

$$(3.6.27) \quad R(0) = \sum_{m=0}^{n-1} A^m S^2 A^{*m} + A^n R(0) A^{*n}.$$

This equation leads to (3.6.14), and T^2 , defined as the limit in (3.6.14), satisfies (3.6.3). Let $\dots, \xi(-1), \xi(0), \dots, \xi$ be mutually independent Gaussian variables satisfying (3.6.1). Then the $x(n)$ defined by (3.6.2) determine the variables of a Gaussian process with non-negative values of n , but a slight modification is needed to obtain an expression defined for all n . To obtain this, it can be supposed that A, T, S are in the forms (3.6.24), (3.6.25). Define \tilde{A} by

$$(3.6.28) \quad \tilde{A}: \begin{pmatrix} I & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & I \end{pmatrix}.$$

Then A is orthogonal and $\tilde{A}T = AT$. If now (3.6.2) is used to define $x(n)$ for all n with A^*T replaced by \tilde{A}^*T , the $x(n)$ process is a t.h.G.M. process with the desired properties.

The properties of the process reversed in time are of some interest. It is easy to see that if n is replaced by $-n$, a t.h.G.M. process remains a t.h.G.M. process. If the original process is non-degenerate, the new transition matrix is $R(0)A^*R(0)^{-1}$. If the transition matrix remains unchanged when n is replaced by $-n$, $R(0)A^*R(0)^{-1} = A$. This is equivalent to the equation $R(n) = R(-n)$.

The simplest generalization of a t.h.G.M. process is the following. Let the chance variables $\{y(n)\}$ determine a t.h.G. process with the property that for some $N > 0$,

$$(3.7.1) \quad E\{\dots, y(n-1); y(n)\} = E\{y(n-N), \dots, y(n-1); y(n)\},$$

with probability 1. If $N = 1$, the process is a t.h.G.M. process. This type process will be called a t.h.G.M._N process. To avoid notational complications only the one-dimensional case will be considered. The right hand side of (3.7.1) is a linear combination of the variables $y(n-N), \dots, y(n-1)$. The variables thus satisfy a difference equation of the form

$$(3.7.2) \quad y(n) - a_1 y(n-1) - \dots - a_N y(n-N) = \eta(n)$$

generalizing (3.6.9), where $\eta(n)$ is independent of the chance variables $\dots, y(n-2), y(n-1)$. The $\{\eta(n)\}$ are mutually independent chance variables with zero means and dispersions independent of n . Equation (3.7.2) leads to

$$(3.7.3) \quad y(n) - a_1^{(n-m)} y(m-1) - \dots - a_N^{(n-m)} y(m-N) = \eta^{(n-m)}(n) \quad (m \leq n)$$

where $\eta^{(n-m)}(n)$ has zero mean and is independent of the chance variables $\dots, y(m-2), y(m-1)$. Hence

$$(3.7.4) \quad E\{\dots, y(m-1); y(n)\} = E\{y(m-N), \dots, y(m-1); y(n)\},$$

$$m \leq n.$$

The difference equation (3.7.2) has been studied in some detail in the past.¹⁴ We shall use an approach which adds insight into the structure of the solution and which clarifies the place of the solution in the general theory of t.h.G. processes. This approach is in terms of N -dimensional t.h.G.M. processes. Define the variables $\{x(n)\}$ of an N -dimensional process by

$$(3.7.5) \quad x_j(n) = y(n+j), \quad n = 0, \pm 1, \dots, \quad j = 1, \dots, N.$$

The $x(n)$ process is evidently a t.h.G.M. process. If the index N of the $y(n)$ process is the minimum for which (3.7.1) is true, the corresponding $x(n)$ process will be non-degenerate. Then the transition matrix A is uniquely determined, and is evidently

$$(3.7.6) \quad A: \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \\ a_N & \cdot & \cdot & \cdot & \cdot & \cdot & a_1 \end{pmatrix} \quad a_1 \neq 0.$$

The matrix S , measuring the dispersion of the prediction $Ax(n-1)$ of $x(n)$, has the form

$$(3.7.7) \quad S: \begin{pmatrix} 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & s \end{pmatrix}.$$

The characteristic equation of A is simply

$$(3.7.8) \quad \alpha^N - a_1\alpha^{N-1} - \dots - a_N = 0.$$

The matrix A has only a single characteristic vector corresponding to each characteristic value λ , the vector $(1, \lambda, \dots, \lambda^{N-1})$. Hence if λ is a multiple root of (3.7.8), it does not correspond to a simple elementary divisor. Therefore, according to Theorem 3.6, all roots of (3.7.8) of modulus 1 must be simple roots. It will be proved below that either no roots have modulus 1 or all roots have modulus 1.

If an N -dimensional non-degenerate t.h.G.M. process is given whose transition matrix A and dispersion matrix S have the forms (3.7.6) and (3.7.7) respectively,

$$x_j(n) - x_{j+1}(n-1) = 0$$

with probability 1, for $j < N$. Then a $y(n)$ process can be defined unambiguously by (3.7.5). Since for fixed j , $x_j(n)$ determines a one-dimensional t.h.G. process, the $y(n)$ process is a t.h.G. process, and (3.7.1) is obviously true, with N minimal if A is non-singular.

Case 1. $S = 0$ (deterministic case). In this case the $x(n)$ process is deterministic, and the $y(n)$ process satisfies the equation

$$(3.7.2') \quad y(n) = a_1 y(n-1) + \dots + a_N y(n-N).$$

¹⁴ Cf. for example H. Wold, *A Study in the Analysis Of Stationery Time Series*, Uppsala, 1938.

All the roots of (3.7.8) are simple roots, of modulus 1. Since $S = 0$,

$$(3.7.9) \quad x(n) = A^n T \xi \quad n = 0, \pm 1, \dots$$

and therefore

$$(3.7.10) \quad y(n) = x_1(n-1) = \sum_{j=1}^N (A^{n-1} T)_{1j} \xi_j \quad n = 0, \pm 1, \dots$$

Using either the well known form of the solution of the N th order difference equation (3.7.2') or of the powers of an orthogonal matrix, it follows that

$$(3.7.11) \quad y(n) = \sum_{j=1}^N (\eta_j \cos n\theta_j + \zeta_j \sin n\theta_j)$$

where the η_j and ζ_j are (one-dimensional) Gaussian variables, and

$$\{\cos \theta_j + i \sin \theta_j\}$$

are the N distinct characteristic values of A , that is the roots of (3.7.8).

Case 2. $S \neq 0$ (non-deterministic case). In this case it will now be shown that the $x(n)$ process can have no deterministic factors: that is that the roots of (3.7.8) all have modulus less than 1. In fact let β be a root of (3.7.8), corresponding to the characteristic vector z of A^* :

$$(3.7.12) \quad \begin{aligned} z &= (a_N \beta^{N-1}, a_N \beta^{N-2} + a_{N-1} \beta^{N-1}, \dots, a_N + a_{N-1} \beta + \dots + a_1 \beta^{N-1}) \\ &= (a_N \beta^{N-1}, \dots, \beta^N). \end{aligned}$$

Then using (3.6.4),

$$(3.7.13) \quad \begin{aligned} (R(0)z, z) &= (AR(0)A^*z, z) + (S^2z, z) \\ &= (R(0)A^*z, A^*z) + (Sz, Sz) \\ &= |\beta|^2 (R(0)z, z) + s^2 |\beta|^{2N}. \end{aligned}$$

Hence $|\beta|$ cannot be 1, and the $x(n)$ process can have no deterministic factors. Equation (3.6.2) becomes

$$(3.7.14) \quad x(n) = \sum_{m=0}^{\infty} A^m S \xi(n-m)$$

which leads to

$$(3.7.15) \quad y(n) = \sum_{j=1}^N \sum_{m=0}^{\infty} (A^m S)_{1j} \xi_j(n-m-1) = s \sum_{m=0}^{\infty} (A^m)_{1N} \xi_N(n-m-1).$$

According to Theorem 3.6 the only restriction on the coefficients a_1, \dots, a_N in the two cases $S = 0$, $S \neq 0$, are respectively that equation (3.7.8) has N distinct roots of modulus 1 and all roots of modulus less than 1. Hence (3.7.10) and (3.7.15) furnish (with the stated restrictions on A) the most general t.h.G.M.N. processes.

It was shown in Theorem 3.6 that if $R(n)$ is the correlation function of a t.h.G.M. process, $R(n)$ can be expressed in the form (3.6.5), where A is some suitably chosen matrix. Conversely if the correlation function of a t.h.G. process has the form (3.6.5), the process is a t.h.G.M. process since $x(n+1) - Ax(n)$ is then orthogonal to (and therefore independent of) the variables $\dots, x(n-1), x(n)$. (This fact implies the truth of (3.1.1)). The characterization of t.h.G.M. processes in terms of their correlation functions is thus easily solved. The following theorems characterize one-dimensional t.h.G.M._N processes from various points of view. It will be convenient, and also intrinsically interesting to treat at the same time a slightly larger class of processes: the class of *component processes of t.h.G.M. processes*. A one-dimensional t.h.G. process with variables $\{x_1(n)\}$ will be called a component process of an N -dimensional t.h.G.M. process if there are $N-1$ t.h.G. processes with variables $\{x_2(n)\}, \dots, \{x_N(n)\}$ such that the N -dimensional process with variables $\{x_1(n), \dots, x_N(n)\}$ is a t.h.G.M. process. If the variables $\{x(n)\}$ determine an N -dimensional t.h.G.M. process, the t.h.G. processes determined by $\{x_1(n)\}, \dots, \{x_N(n)\}$ will be called its N component processes. If an $x(n)$ process is not of type $M(0)$ and is a component process of an N -dimensional t.h.G.M. process, it is a component process of a non-degenerate N_1 -dimensional t.h.G.M. process, for some $N_1 \leq N$. It has already been seen that one-dimensional t.h.G.M._N processes are component processes of N -dimensional t.h.G.M. processes.

THEOREM 3.8. *Let $\dots, x(0), x(1), \dots$ be the variables of a one-dimensional t.h.G. process. The process is a component process of an N -dimensional t.h.G.M. process if and only if the chance variables*

$$(3.8.1) \quad x(0), E\{\dots, x(-1), x(0); x(n)\}, \quad n = 1, 2, \dots$$

are linearly dependent on the first N .

Suppose that the $x(n)$ process is a component process of an N -dimensional $y(n)$ process: $x(n) = y_1(n)$, with correlation function $R_y(n)$ and transition matrix A . Since A satisfies its characteristic equation

$$(3.8.2) \quad \det |\alpha I - A| = \alpha^N - a_1 \alpha^{N-1} - \dots - a_N = 0,$$

it follows from (3.6.2) that if $\eta(n+N)$ is defined by

$$(3.8.3) \quad y(n+N) - a_1 y(n+N-1) - \dots - a_N y(n) = \eta(n+N)$$

then $\eta(n+N)$ is independent of $\dots, y(n-1), y(n)$. Then

$$(3.8.4) \quad x(n+N) - a_1 x(n+N-1) - \dots - a_N x(n) = \eta_1(n+N)$$

where $\eta_1(n+N)$ is independent of the chance variables $\dots, x(n-1), x(n)$. Equation (3.8.4) leads to

$$(3.8.5) \quad x(n+N+\nu) - a_1^{(\nu)} x(n+N-1) - \dots - a_N^{(\nu)} x(n) = \eta_1^{(\nu)}(n+N+\nu)$$

It is easily verified (using the fact that $R_q(n) = R_q(-n) = \overline{R_q(n)}$) that if α is a root of the equation

$$(3.9.2) \quad \sum_{n=-(N-1)}^{N-1} R_q(n) z^n = 0,$$

then $\bar{\alpha}$, $1/\alpha$, $1/\bar{\alpha}$ are also roots, of the same multiplicity. Moreover if $|\alpha| = 1$, α is a root of even multiplicity, since the sum in (3.9.2) is real and non-negative when $|z| = 1$. When $|z| = 1$,

$$(3.9.3) \quad |\alpha| |(z - \alpha)(z - 1/\bar{\alpha})| = |z - \alpha|^2.$$

Hence $G'_q(\lambda)$ can be written in the following simple form:

$$(3.9.4) \quad G'_q(\lambda) = |\beta_0 e^{i(N-1)\lambda} + \beta_1 e^{i(N-2)\lambda} + \dots + \beta_{N-1}|^2$$

where the roots of the indicated polynomial have modulus at most 1, and the coefficients are real.

THEOREM 3.9. Let $\dots, x(0), x(1), \dots$ be the variables of a one-dimensional t.h.G. process. The process is a component process of a finite-dimensional t.h.G.M. process if and only if the complex spectral function is the sum of the integral of the square of the absolute value of a rational function of $e^{i\lambda}$ with real coefficients, and of a monotone non-decreasing function increasing only in a finite number of jumps¹⁵. Specifically:

(i) The process is a component process of an N -dimensional t.h.G.M. process if and only if the complex spectral function has the form

$$(3.9.5) \quad G(\lambda) = \int_{-\pi}^{\lambda} \frac{|\beta_0 e^{i(N-1)\lambda} + \dots + \beta_{N-1}|^2}{|\alpha_0 e^{iN\lambda} + \dots + \alpha_N|^2} d\lambda + \hat{G}(\lambda)$$

where

- (a) $\hat{G}(\lambda)$ is a monotone non-decreasing function satisfying (1.3.3), increasing only by jumps, at no more than N points;
- (b) the denominator of the integrand vanishes at every discontinuity of $G(\lambda)$, and the numerator vanishes at every zero of the denominator, to at least the same order;
- (c) the coefficients $\alpha_0, \dots, \alpha_N, \beta_0, \dots, \beta_{N-1}$ are real, $\alpha_0 \neq 0, \beta_0 \neq 0$ unless the integrand vanishes identically, and the roots of the polynomials in the integrand have modulus less than or equal to 1.

The integral vanishes identically if and only if the $x(n)$ process is a component process of an N -dimensional deterministic process, and $\hat{G}(\lambda)$ vanishes identically if and only if the variables $x(n)$ vanish identically or the $x(n)$ process is a component process of an N -dimensional t.h.G.M. process with no deterministic factor.

(ii) The process is a t.h.G.M._N process (deterministic case) if and only if the

¹⁵ It is easily seen that the first term of the two can also be described simply as the integral of a rational function of $e^{i\lambda}$, which is non-negative for λ real and is an even function of λ like all complex spectral density functions,

complex spectral function $G(\lambda) = \hat{G}(\lambda)$ is a monotone non-decreasing function satisfying (1.3.3) increasing only in jumps, at no more than N points; (non-deterministic case) if and only if the complex spectral function has the form

$$(3.9.6) \quad G(\lambda) = \int_{-\pi}^{\lambda} \frac{d\lambda}{|\alpha_0 e^{iN\lambda} + \dots + \alpha_N|^2}$$

where $\alpha_0, \dots, \alpha_N$ are real and $\alpha_0 \neq 0$.

Proof of (i). If the $x(n)$ process is a one-dimensional component of an N -dimensional t.h.G.M. process, it has already been seen that for properly chosen real numbers a_1, \dots, a_N , (3.8.4) is true, where $\eta_1(n+N)$ is independent of the chance variables $\dots, x(n-1), x(n)$. Equation (3.8.2) can be assumed to have all its roots of modulus less than or equal to 1. It follows from (3.8.4) that $\eta_1(n)$ is independent of $\eta_1(m)$ if $|n-m| \geq N$. The complex spectral function of the $\eta_1(n)$ process is therefore continuous, with derivative given by (3.9.4). It will be no restriction to assume that $b_0 \neq 0$ unless the derivative vanishes identically. According to (1.3.6), if $G(\lambda)$ is the complex spectral function of the $x(n)$ process,

$$(3.9.7) \quad \begin{aligned} E\{\eta_1(0)\eta_1(n)\} &= \int_{-\pi}^{\pi} e^{i\lambda n} |b_0 e^{i(N-1)\lambda} + \dots + b_{N-1}|^2 d\lambda \\ &= \int_{-\pi}^{\pi} e^{i n \lambda} |e^{i N \lambda} - a_1 e^{i(N-1)\lambda} - \dots - a_N|^2 dG(\lambda). \end{aligned}$$

Hence if $\hat{G}(\lambda)$ is the jump function of $G(\lambda)$ ($\hat{G}(-\pi) = 0$, and $\hat{G}(\lambda)$ is constant except for jumps at the same points as those of $G(\lambda)$, and of the same magnitude),

$$(3.9.8) \quad \begin{aligned} &\int_{-\pi}^{\lambda} |b_0 e^{i(N-1)\lambda} + \dots + b_{N-1}|^2 d\lambda \\ &= \int_{-\pi}^{\lambda} |e^{i N \lambda} - \dots - a_N|^2 d[G(\lambda) - \hat{G}(\lambda)] \\ &\quad + \int_{-\pi}^{\lambda} |e^{i N \lambda} - \dots - a_N|^2 d\hat{G}(\lambda). \end{aligned}$$

Since the first two integrals are continuous in λ , the last must be continuous also. Hence the last integrand must vanish at every discontinuity of $\hat{G}(\lambda)$, that is at every discontinuity of $G(\lambda)$, and the last integral vanishes identically. It follows that

$$(3.9.9) \quad G(\lambda) - \hat{G}(\lambda) = \int_{-\pi}^{\lambda} \frac{|b_0 e^{i(N-1)\lambda} + \dots + b_{N-1}|^2}{|e^{i N \lambda} - \dots - a_N|^2} d\lambda$$

where the numerator vanishes at each zero of the denominator, with the same or greater multiplicity. Since the denominator vanishes at each discontinuity of $G(\lambda)$, there can be at most N discontinuities. If the N -dimensional process is a deterministic process, it can be assumed that all the roots of equation (3.8.2)

have modulus 1, that is that the denominator and hence also the numerator in (3.9.9) have N roots. This can be true only if the numerator vanishes identically: $G(\lambda) \equiv \hat{G}(\lambda)$. If the N -dimensional process has no deterministic factor, it can be assumed that all the roots of equation (3.8.2) have modulus less than 1. Then $G(\lambda)$ can have no discontinuities: $\hat{G}(\lambda) \equiv 0$.

Conversely if $G(\lambda)$ has the form described in Theorem 3.9 (i), $G(\lambda)$ can be assumed in the form (3.9.9) with real coefficients in numerator and denominator and the stated relations between the jumps of $G(\lambda)$ and the zeros of the numerator and denominator in the integrand. (If the integrand vanishes identically and if $G(\lambda)$ has N discontinuities, a_1, \dots, a_N can be chosen as those numbers making the polynomial

$$e^{iN\lambda} - a_1 e^{i(N-1)\lambda} - \dots - a_N$$

vanish at the discontinuities of $G(\lambda)$.) Then

$$\begin{aligned} R(n+N) - a_1 R(n+N-1) - \dots - a_N R(n) \\ (3.9.10) \quad = \int_{-\pi}^{\pi} e^{i\lambda(n+1)} \frac{[b_0 + \dots + b_{N-1} e^{i(N-1)\lambda}][b_0 e^{i(N-1)\lambda} + \dots + b_{N-1}] d\lambda}{1 - a_1 e^{i\lambda} - \dots - a_N e^{iN\lambda}} \\ + \int_{-\pi}^{\pi} e^{i\lambda n} [e^{iN\lambda} - \dots - a_N] d\hat{G}(\lambda). \end{aligned}$$

The last integral vanishes since the bracket vanishes at every jump of $\hat{G}(\lambda)$. The denominator in the first integrand is the value on $|z| = 1$ of a polynomial all of whose roots are outside $|z| = 1$, or on $|z| = 1$. Any zero on $|z| = 1$ corresponds to one of the numerator at the same point. The integral therefore vanishes if $n \geq 0$ (Cauchy Integral Theorem):

$$(3.9.11) \quad R(n+N) - a_1 R(n+N-1) - \dots - a_N R(n) = 0, \quad n \geq 0.$$

This equation implies that

$$(3.9.12) \quad x(n+N) - a_1 x(n+N-1) - \dots - a_N x(n)$$

is independent of the chance variables $\dots, x(n-1), x(n)$, that is that (3.8.4) is true, where $\eta_1(n+N)$ has the stated properties. It has already been seen in the proof of Theorem 3.8 that this implies (3.8.6) and that this in turn implies that the process is a component process of an N -dimensional t.h.G.M. process whose transition matrix A has characteristic equation (3.8.2). In particular if $G(\lambda) \equiv \hat{G}(\lambda)$, the roots of the characteristic equation are of modulus 1, so that the N -dimensional process must be deterministic. If $\hat{G}(\lambda) \equiv 0$, the $x(n)$ process is a component process of an N -dimensional process whose transition matrix A has only characteristic values of modulus less than 1. This N -dimensional process can have no deterministic factors other than one or more of type $M(0)$. If these exist, (and if the $x(n)$ process is not of type $M(0)$) they can be replaced by non-degenerate factors of type M , to obtain an N -dimensional process with no deterministic factor, having the $x(n)$ process as a component process.

Proof of (ii). If the $x(n)$ process is a t.h.G.M._N process, (3.8.4) is true with $\eta_1(m)$ independent of $\eta_1(n)$ if $m \neq n$. The discussion in (i) is therefore simplified by the fact that the numerator in (3.9.9) is constant. If this constant is 0, the spectral function is a function of jumps: $G(\lambda) \equiv \hat{G}(\lambda)$. If this constant is not 0, the denominator in (3.9.9) does not vanish, and $\hat{G}(\lambda)$ therefore vanishes identically. The converse is proved as in (i).

THEOREM 3.10. (i) If a_1, \dots, a_N are real numbers, there is a one-dimensional t.h.G. process not of type $M(0)$ with correlation function $R(n)$ satisfying

$$(3.10.1) \quad R(n+N) - a_1 R(n+N-1) - \dots - a_N R(n) = 0$$

for $n \geq 0$ if and only if the equation

$$(3.10.2) \quad \alpha^N - a_1 \alpha^{N-1} - \dots - a_N = 0$$

has at least one root of modulus less than or equal to 1.

Let $\dots, x(0), x(1), \dots$ be the variables of a one-dimensional t.h.G. process not of type $M(0)$.

(ii) This process is a component process of an N -dimensional t.h.G.M. process if and only if the correlation function $R(n)$ satisfies an N th order linear difference equation (3.10.1) for $n \geq 0$.

(iii) The process is a t.h.G.M._N process if and only if the difference equation (3.10.1) is true for $n \geq -(N-1)$. In this case the vectors $\{x(n), \dots, x(n+N-1)\}$ determine an N -dimensional t.h.G.M. process.

(iv) Equation (3.10.1) is satisfied for $n \geq -N$ if and only if

$$(3.10.3) \quad x(n+N) - a_1 x(n+N-1) - \dots - a_N x(n) = 0, \quad n = 0, \pm 1, \dots$$

Proof of (ii), (iii), (iv). Let the $x(n)$ process be a component process of an N -dimensional t.h.G.M. $y(n)$ process with correlation function $R_v(n)$: $x(n) = y_1(n)$, and transition matrix A . Since A satisfies its characteristic equation (3.8.2), it follows from (3.6.5) that

$$(3.10.4) \quad R_v(n+N) - a_1 R_v(n+N-1) - \dots - a_N R_v(n) = 0, \quad n \geq 0.$$

Then $R(n) = (R_v(n))_{11}$ satisfies this same difference equation. Conversely if (3.10.1) is true for $n \geq 0$, it has already been proved in the course of the proof of Theorem 3.9 that the $x(n)$ process is a component process of an N -dimensional t.h.G.M. process. This finishes the proof of (ii). Parts (iii) and (iv) are proved similarly.

Proof of (i). According to (ii), if there is a one-dimensional t.h.G. process whose correlation function $R(n)$ satisfies (3.10.1) for $n \geq 0$, the process is a component process of an N -dimensional t.h.G.M. process whose transition matrix A has (3.10.2) as characteristic equation. Since A has at least one characteristic value of modulus less than or equal to 1, (unless the $x(n)$ process is of type $M(0)$), (3.10.2) must have at least one root of modulus less than or equal to 1. Conversely if (3.10.2) has at least one such root, there is a real N -dimensional matrix A whose characteristic equation is (3.10.2), and which has simple ele-

mentary divisors. According to Theorem 3.6 (ii), A is then the transition matrix of some t.h.G.M. process. The correlation function of this process and hence that of each component process satisfies (3.10.1) for $n \geq 0$.

THEOREM 3.11. (i) If a_1, \dots, a_N are real numbers, there is a one-dimensional t.h.G. process not of type $M(0)$ satisfying

$$(3.11.1) \quad x(n+N) - a_1 x(n+N-1) - \dots - a_N x(n) = \eta(n+N),$$

$$n = 0, \pm 1, \dots$$

with $\eta(m), \eta(n)$ independent for $|m-n| \geq N$ if and only if (3.10.2) has at least one root of modulus less than or equal to 1.

Let $\dots, x(0), x(1), \dots$ be the variables of a one-dimensional t.h.G. process.

(ii) This process is a component process of an N -dimensional t.h.G.M. process if and only if (3.11.1) is true with $\eta(m), \eta(n)$ independent for $|m-n| \geq N$. In this case $\eta(n+N)$ will be independent of the chance variables $\dots, x(n-1), x(n)$.

(iii) The process is a t.h.G.M. process if and only if in addition to the condition in (ii), $\eta(n)$ is independent of the chance variables $\dots, x(n-2), x(n-1)$: deterministic case if $\eta(n) = 0$ with probability 1, nondeterministic case otherwise.

Since this theorem follows readily from the preceding theorems, the proof will be omitted.

The problem of predicting $x(n)$ in terms of $\dots, x(n-2), x(n-1)$ is trivial (theoretically at least) for t.h.G.M. processes. In fact these were defined as those processes for which the solution of the prediction problem is simply a linear combination $\sum_{j=1}^N a_j x(n-j)$ of the preceding N variables. The solution will now be given for the more general class of component processes of N -dimensional t.h.G.M. processes, processes which have been described from several points of view in the preceding theorems.

The prediction problem for component processes of N -dimensional t.h.G.M. processes will be put into a more general setting. If the one-dimensional chance variables $\{x(n)\}$ determine a t.h.G. process, with correlation function $R(n)$, the problem of finding $E\{\dots, x(n-2), x(n-1); x(n)\}$ is that of finding a series $\sum_{m=1}^{\infty} \gamma_m x(n-m)^{16}$ such that

$$(3.12.1) \quad x(n) - \sum_{m=1}^{\infty} \gamma_m x(n-m)$$

is uncorrelated with every $x(n-\nu)$ ($\nu > 0$):

$$(3.12.2) \quad R(\nu) - \sum_{m=1}^{\infty} \gamma_m R(\nu-m) = 0, \quad \nu > 0.$$

¹⁶ We are neglecting all convergence difficulties. They become trivial for the applications to be made below.

If the complex spectral function is $G(\lambda)$, (3.12.2) becomes

$$(3.12.3) \quad \int_{-\pi}^{\pi} e^{i\nu\lambda} \left\{ 1 - \sum_{m=1}^{\infty} \gamma_m e^{-im\lambda} \right\} dG(\lambda) = 0, \quad \nu = 0.$$

Let $G(\lambda)$ be the integral of its derivative $G'(\lambda)$, that is let $G(\lambda)$ be absolutely continuous. According to (3.12.3) the problem reduces to that of finding a function

$$(3.12.4) \quad f(z) = 1 - \frac{\gamma_1}{z} - \frac{\gamma_2}{z^2} - \dots \quad z = e^{i\lambda}$$

such that $f(z)G'$ is of power series type, a function corresponding to an expansion in non-negative powers of z . The dispersion of the error of the prediction is

$$(3.12.5) \quad E \left\{ \left[x(n) - \sum_{m=1}^{\infty} \gamma_m x(n-m) \right]^2 \right\} = \int_{-\pi}^{\pi} \left| e^{in\lambda} - \sum_{m=1}^{\infty} \gamma_m e^{i(n-m)\lambda} \right|^2 dG(\lambda) \\ = \int_{-\pi}^{\pi} |f|^2 dG(\lambda).$$

In particular if the $x(n)$ process is a component process of an N -dimensional t.h.G.M. process, $G(\lambda)$ is given by (3.9.5). It will be supposed throughout the following that $\hat{G} \equiv 0$. Then

$$(3.12.6) \quad G'(\lambda) = \frac{z(\beta_0 z^{N-1} + \dots + \beta_{N-1})(\beta_0 + \dots + \beta_{N-1} z^{N-1})}{(\alpha_1 z^N + \dots + \alpha_N)(\alpha_0 + \dots + \alpha_N z^N)}, \quad z = e^{i\lambda}.$$

In this case $f \equiv 1$ if $G' \equiv 0$, and otherwise f is given by

$$(3.12.7) \quad f(z) = \frac{\beta_0(\alpha_0 z^N + \dots + \alpha_N)}{\alpha_0 z(\beta_0 z^{N-1} + \dots + \beta_{N-1})}$$

so that

$$(3.12.8) \quad |f(z)|^2 = \frac{\beta_0^2}{\alpha_0^2 G'}.$$

The dispersion of the prediction error is $R(0)$ if $G' \equiv 0$ and otherwise is $2\pi\beta_0^2/\alpha_0^2$. The prediction formula for $x(n)$ in terms of the variables $\dots, x(n-\nu-1), x(n-\nu)$ has now been derived for $\nu = 1$, for the chance variables under discussion in this section. The solution for general ν is easily obtained.

As $\nu \rightarrow \infty$, the prediction converges with probability 1, according to the corollary to Theorem 1.2. If the process is a component process of an N -dimensional t.h.G.M. process, and if $\hat{G} \equiv 0$ in (3.9.5), the limit is 0. That is, in this case, the best predicted value of $x(n)$ in terms of the distant past is near $E\{x(n)\} = 0$, the same predicted value which would be used with no knowledge of the past.

4. Processes whose parameter t varies continuously

The basic process in terms of which t.h.G.M. processes without deterministic factors were expressed in section 3 was a process whose variables $\{\xi(n)\}$ were Gaussian, with

$$(4.1.1) \quad E\{\xi(n)\} = 0, \quad E\{\xi(m) \cdot \xi(n)\} = \delta_{m,n}I.$$

The corresponding process in the continuous parameter case is not obtained by replacing the integral parameters m, n in (4.1.1) by continuous parameters. In fact the process so defined does not satisfy any useful continuity conditions. In the present discussion, sums like $\sum_m A_m \xi(m)$ will be replaced by Stieltjes

integrals $\int A(t) d\xi(t)$, and $d\xi(t)$ thus will correspond to $\xi(n)$. The $\xi(t)$ process is defined as follows. For any $t_1 < \dots < t_n$, the chance variables

$$\xi(t_2) - \xi(t_1), \dots, \xi(t_n) - \xi(t_{n-1})$$

are mutually independent N -dimensional Gaussian chance variables, and if $s < t$,

$$(4.1.2) \quad E\{\xi(t) - \xi(s)\} = 0, \quad E\{[\xi(t) - \xi(s)] \cdot [\xi(t) - \xi(s)]\} = (t - s)I.$$

This process, called simply a ξ -process below has been discussed in great detail by Bachelier, Wiener and Lévy. The function $\xi(t)$, considered as a function of t is known to be continuous with probability 1.¹⁷ The derivative $\xi'(t)$ does not exist, since $E\{[\xi_s(t+h) - \xi_s(t)]^2\}$ is proportional to h , whereas this mean would be proportional to h^2 if $\xi'(t)$ existed. In fact it has been shown that $\xi(t)$ is (with probability 1) not even of bounded variation in any finite interval. However, if $f(t)$ is a function defined and continuous for $a \leq t \leq b$ (where a or b or both may be infinite), the integral

$$(4.1.3) \quad \int_a^b f(t) d\xi(t)$$

can be defined as the limit in the mean of the usual Stieltjes sum. If $f(t)$ has a continuous derivative, the integral in (4.1.3) can be evaluated by integration by parts:

$$(4.1.4) \quad \int_a^b f(t) d\xi(t) = f(b)\xi(b) - f(a)\xi(a) - \int_a^b \xi(t)f'(t) dt.$$

Integrals of the following type will be used below:

$$(4.1.5) \quad \begin{aligned} y(t) = \int_a^t f(t - \tau) d\xi(\tau) &= f(0)\xi(t) - f(t - a)\xi(a) \\ &\quad + \int_a^t \xi(\tau)f'(t - \tau) d\tau \end{aligned}$$

¹⁷ Paley and Wiener, "Fourier transforms in the complex domain," *Am. Math. Soc. Colloq. Pub.*, Vol. 19, p. 148.

where $f(t)$ is continuous and has two continuous derivatives. It is then evident that $y(t)$ is continuous, but that $y'(t)$ exists if and only if $f(0) = 0$. If $f(0) = 0$, $y'(t)$ is given by

$$(4.1.6) \quad y'(t) = \int_a^t f'(t - \tau) d\xi(\tau).$$

A more general process will also come into the discussion below, and will be called a ζ -process. The chance variables $\{\zeta(t)\}$ of a ζ -process are Gaussian, and have the same independence property as the variables of a ξ -process. The second equation of (4.1.2) is dropped, so that (4.1.2) is replaced by

$$(4.1.7) \quad E\{\zeta(t) - \zeta(0)\} = 0, \quad E\{[\zeta(t) - \zeta(0)] \cdot [\zeta(t) - \zeta(0)]\} = D(t),$$

where the symmetric and non-negative definite matrix $D(t)$ will sometimes be supposed to have special properties, such as continuity in t , etc. The independence property of the ζ -process implies that

$$(4.1.8) \quad E\{[\zeta(t) - \zeta(s)] \cdot [\zeta(t) - \zeta(s)]\} = D(t) - D(s).$$

Hence $D'(t)$ (if this derivative exists) is symmetric and non-negative definite.

THEOREM 4.1. *If the dispersion matrix $D(t)$ of a ζ -process is continuous, the functions $\{\zeta(t)\}$ are continuous in t , with probability 1.*

The component processes of a ζ -process with a continuous dispersion function are also ζ -processes with continuous dispersion functions. Hence it will be sufficient to prove the theorem in the one-dimensional case. In this case $D(t)$ is non-negative and monotone non-decreasing, according to (4.1.7) and (4.1.8). It can be supposed that $D(t)$ does not vanish identically. Let $D_1(t)$ be an inverse function of $D(t)$: $D[D_1(t)] = t$. Then $\xi(t) = \zeta[D_1(t)]$ defines a ξ -process, and the continuity of $\xi(t)$ implies that of $\zeta(t)$.

The integrals of type (4.1.3) are defined for ζ -processes as for ξ -processes, and satisfy the equations

$$(4.1.9) \quad \begin{aligned} E \left\{ \int_a^b f(t) d\zeta(t) \right\} &= 0 \\ E \left\{ \int_a^b f(t) d\zeta(t) \cdot \int_a^b g(t) d\zeta(t) \right\} &= \int_a^b f(t)g(t)D'(t) dt \\ E \left\{ \int_a^b A(t) d\zeta(t) \cdot \int_a^b B(t) d\zeta(t) \right\} &= \int_a^b A(t)D'(t)B(t)^* dt, \end{aligned}$$

where f, g are numerically valued functions and A, B are matrix functions.¹⁸

The ζ -processes lie at the basis of t.h.G. processes. To every t.h.G. process (discrete parameter) with variables $\{x(n)\}$ correspond two one-dimensional ζ -processes with variables $\{\zeta_1(t)\}, \{\zeta_2(t)\}$ such that

$$(4.1.10) \quad x(n) = \int_{-\pi}^{\pi} \cos n\lambda d\zeta_1(\lambda) + \sin n\lambda d\zeta_2(\lambda)$$

¹⁸ These equations are easily proved using the fact that each integral can be approximated by the usual Riemann-Stieltjes sums.

where the two ζ -processes are mutually independent in the sense that every $\zeta_1(\lambda_1)$ is independent of every $\zeta_2(\lambda_2)$ and where, if $G(\lambda)$ is the complex spectral function of the process,

$$(4.1.11) \quad E\{\zeta_1(\lambda)^2\} = G(\lambda).$$

In the continuous parameter case (4.1.11) becomes

$$(4.1.12) \quad x(t) = \int_{-\infty}^{\infty} \cos t\lambda \, d\zeta_1(\lambda) + \sin t\lambda \, d\zeta_2(\lambda).^{19}$$

This theorem of Cramér shows that $x(n)$, or $x(t)$ as the case may be, is the limit of a sum of sines and cosines, with Gaussian chance variables as coefficients. The dispersion of each coefficient, which measures the intensity of the corresponding periodic term of the sum, is determined by the spectral function of the process. In particular, if the spectral function $F(\lambda)$ is the integral of its derivative $F'(\lambda)$, each integrand involving $d\zeta_1(\lambda)$ in the above equations can be replaced by one involving $\sqrt{F'(\lambda)}d\xi_1(\lambda)$ where $\xi_1(\lambda)$ is the variable of a ξ -process. Thus in many important cases the processes can be written in a simple way in terms of ξ -processes.

It will be shown below that every t.h.G. M. process can be represented as the direct product of factors of certain types. The deterministic types have already been catalogued: $M(0)$, $M(1)$, $M(e^{i\theta})$. The standard non-deterministic type, as in the discrete parameter case, will be called type M .

M Let $\{\xi(t)\}$ be the chance variables of an N -dimensional ξ -process, as described above. Let Q be an N -dimensional square matrix, and let S be an N -dimensional symmetric non-negative definite matrix. Define $x(t)$ by

$$(4.2.1) \quad x(t) = \int_0^{\infty} e^{tQ} S \, d\xi(t-s) = \int_{-\infty}^t e^{(t-s)Q} S \, d\xi(s)$$

where it is supposed that the improper integrals converge with probability 1. There will be convergence, for example, if Q has only characteristic values with negative real parts so that the elements in the matrix e^{sQ} go to 0 exponentially as $s \rightarrow \infty$. (Cf section 1.) It will be shown below that it is no restriction to assume that Q has this character. The $x(t)$ process is evidently a t.h.G. process. If $u < t$, the chance variable

$$(4.2.2) \quad x(t) - e^{(t-u)Q} x(u) = e^{tQ} \int_u^t e^{-sQ} S \, d\xi(s)$$

is independent of $x(u)$ for $v \leq u$, since $x(v)$ is expressed in terms of $\xi(s)$ for $s \leq v$. Therefore the $x(t)$ process is a Markoff process with transition matrix $A(t) = e^{tQ}$:

$$(4.2.3) \quad E\{x(v), v \leq u; x(t)\} = e^{(t-u)Q} x(u), \quad u < t.$$

¹⁹ H. Cramér, *Arkiv For Matematik, Astronomi och Fysik*, Vol. 28B, No. 12, pp. 1-17. Cramér only discusses the continuous parameter case, but the other requires no change of method. He allows complex-valued ξ -processes, in terms of which (4.1.10) and (4.1.12) assume a more elegant form.

A process defined in this way will be called a process of type M . A change of variable $y(t) = Bx(t)$ leads to a process of the same type:

$$(4.2.4) \quad y(t) = \int_0^\infty e^{tBQB^{-1}} BS d\xi(t-s).$$

The matrix Q goes into BQB^{-1} and if S_1O is the polar form of BS , where S is symmetric and non-negative definite and O is orthogonal, S goes into S_1 . (We are using the fact that $O\xi(t)$ defines a second ξ -process.) The correlation function of a process of type M is easily calculated:

$$(4.2.5) \quad R(0) = \int_0^\infty e^{sQ} S^2 e^{sQ^*} ds, \quad R(t) = R(0)e^{tQ^*}.$$

The only condition imposed on Q is that the improper integrals in (4.2.1) converge. This condition is easily seen to be equivalent to the convergence of the integral in (4.2.5). This in turn is equivalent to the condition that

$$(4.2.6) \quad \lim_{t \rightarrow \infty} e^{tQ} S = 0.$$

This condition is certainly satisfied if the characteristic values of Q all have negative real parts, and it can always be assumed that this is so. (Cf. the corresponding discussion of processes of type M in the discrete parameter case.)

The analogues in the continuous parameter case of Theorems 3.4 and 3.5 are true. The proofs are substantially the same as the proofs in the discrete parameter case, and will be omitted.

THEOREM 4.3. (i) Every t.h.G.M. process (continuous parameter) is the direct product of processes of type $M(0)$, $M(1)$, $M(e^{i\theta})$, M .

(ii) If $x(t)$ are the variables of such a process, there is a matrix Q such that $A(t) = e^{tQ}$ is a transition matrix function. There is a ξ -process, a Gaussian variable ξ , independent of the $\xi(t)$, satisfying

$$(4.3.1) \quad E\{\xi\} = 0, \quad E\{\xi \cdot \xi\} = I$$

and symmetric non-negative definite matrices S , T such that

$$(4.3.2) \quad \begin{aligned} x(t) &= \int_0^\infty e^{tQ} S d\xi(t-s) + e^{tQ} T\xi \\ &= \int_0^t e^{(t-s)Q} S d\xi(s) + e^{tQ} T\xi = \int_0^t e^{(t-s)Q} S d\xi(s) + e^{tQ} x(0), \end{aligned}$$

$$(4.3.3) \quad QT^2 + T^2 Q^* = 0,$$

$$(4.3.4) \quad R(0) = \int_0^\infty e^{sQ} S^2 e^{sQ^*} ds + T^2$$

$$(4.3.5) \quad QR(0) + R(0)Q^* = -S^2,$$

where the integrals in (4.3.2) converge with probability 1. The integral and the last term in each pair in (4.3.2) are linear transformations of $x(t)$: (4.3.2) exhibits

in part the decomposition into factor processes described in (i). The correlation function is given by

$$(4.3.6) \quad \begin{aligned} R(t) &= R(0)e^{tQ^*} \\ R(-t) &= e^{tQ}R(0) \end{aligned} \quad t \geq 0.$$

(iii) The matrix Q is uniquely determined if and only if the process is non-degenerate. In any case there is a Q whose characteristic values all have negative or zero real parts and whose characteristic values with zero real parts correspond to simple elementary divisors. The matrix Q furnishes the solution to the prediction problem of the process.

$$(4.3.7) \quad E\{x(s), s \leq t, x(t+u)\} = e^{uQ}x(t), \quad u > 0.$$

The matrix S , which is uniquely determined, measures the dispersion of $x(t)$ about its predicted value:

$$(4.3.8) \quad E\{[x(t+u) - e^{uQ}x(t)]^2\} = R(0) - e^{uQ}R(0)e^{uQ^*} \sim uS^2 \quad (u \rightarrow 0)$$

(iv) Conversely if Q is a matrix with at least one characteristic value with negative real part or with zero real part and corresponding to a simple elementary divisor, e^{tQ} is the transition matrix function of a t.h.G.M. process with $R(0)$ not the null matrix. If all the characteristic values of Q are as just described, e^{tQ} is the transition matrix function of a non-degenerate t.h.G.M. process. If $R(0)$, S , Q are matrices satisfying (4.3.5) with $R(0)$, S symmetric and non-negative definite, there is a t.h.G.M. process whose variables can be written in the form (4.3.2) with the given $R(0)$, S , Q .

The proof of Theorem 4.3 follows closely that of Theorem 3.6, and the details will not be given, except as they differ from those of the earlier proof.

Proof of (i). Suppose that the $\{x(t)\}$ are the variables of a t.h.G.M. process which is non-degenerate. The transition matrix function $A(t)$ is then uniquely determined by (1.5.5). Take the conditional expectation of both sides of (1.5.4) for given $x(0)$:

$$(4.3.9) \quad A(s+t)x(0) = A(t)A(s)x(0) \quad s, t > 0$$

Since the process is non-degenerate,

$$(4.3.10) \quad A(s+t) = A(s)A(t) \quad s, t > 0.$$

According to (1.3.1) and (1.5.5)

$$(4.3.11) \quad \lim_{t \rightarrow 0} R(t) = \lim_{t \rightarrow 0} R(0)A(t)^* = R(0), \quad t > 0.$$

Hence

$$(4.3.12) \quad \lim_{t \rightarrow 0} A(t) = I.$$

It has already been noted that any solution to (4.3.10) under the continuity hypothesis (4.3.12) can be written in the form

$$(4.3.13) \quad A(t) = e^{tQ},$$

where

$$Q = \lim_{t \rightarrow 0} \frac{A(t) - I}{t}.$$

Under a change of variables $y(t) = Bx(t)$, $A(t)$ becomes $BA(t)B^{-1}$ and Q becomes BQB^{-1} . According to Theorem 2.1, if the $x(t)$ process is degenerate, it is the direct product of one or more factors of type $M(0)$ and (perhaps) of a non-degenerate factor. The matrix Q of a factor of type $M(0)$ can be taken as the null matrix. Then the form (4.3.13) is admissible for any t.h.G.M. process, although Q will only be uniquely determined if the process is non-degenerate. Define $\zeta(t)$ by

$$(4.3.14) \quad \zeta(t) = A(t)^{-1}x(t) = e^{-tQ}x(t).$$

Then if $s_1 < t_1 < s_2 < t_2$

$$(4.3.15) \quad E\{\zeta(t_1) - \zeta(s_1)\} = 0, \quad E\{[\zeta(t_2) - \zeta(s_2)] \cdot [\zeta(t_1) - \zeta(s_1)]\} = 0$$

and

$$(4.3.16) \quad D(t) = E[\zeta(t) - \zeta(0)] \cdot [\zeta(t) - \zeta(0)] = e^{-tQ}R(0)e^{-tQ*} - R(0).$$

Hence the $\{\zeta(t)\}$ determine a ζ -process, with dispersion matrix given by (4.3.16). The derivative $D'(t)$ is easily evaluated:

$$(4.3.17) \quad D'(t) = e^{-tQ}[-R(0)Q^* - QR(0)]e^{-tQ*}.$$

Since $D'(t)$ is symmetric and non-negative definite, the bracket also has this property, and there is a non-singular matrix S_1 such that

$$(4.3.18) \quad S_1[-QR(0) - R(0)Q^*]S_1^* = U,$$

where U is in diagonal form, with only 0's and 1's in the main diagonal. Then the integral

$$(4.3.19) \quad \int_0^t S_1 e^{sQ} d\zeta(s)$$

defines a ζ -process with dispersion matrix tU . There is therefore a ξ -process with variables $\{\xi(t)\}$ such that

$$(4.3.20) \quad U\xi(t) = \int_0^t S_1 e^{sQ} d\zeta(s).$$

This equation can be solved for $\xi(t)$ and $x(t)$:

$$(4.3.21) \quad x(t) = e^{tQ}\xi(t) = e^{tQ} \int_0^t e^{-sQ} S_2 U d\zeta(s) + e^{tQ}x(0)$$

where $S_2 = S_1^{-1}$. The matrix $S_2 U$ can be written in the polar form $S O$ where S is symmetric and non-negative definite and O is orthogonal. This S is the S of (4.3.2) etc.

The remainder of the proof follows closely the proof of Theorem 3.6 and will be omitted.

An important class of t.h.G.M. processes which arises frequently in physical applications is obtained in the following way. Let $\{\xi(t)\}$ be the variables of a one-dimensional ξ -process. Consider the formal equation

$$(4.4.1) \quad \frac{d^N y(t)}{dt^N} - a_1 \frac{d^{N-1} y(t)}{dt^{N-1}} - \dots - a_N y(t) = c \xi'(t),$$

where a_1, \dots, a_N, c are constants. This equation cannot be considered precise as it stands, since $\xi'(t)$ does not exist. The problem can however be reformulated as follows: find a $y(t)$ process, where $y', \dots, y^{(N-1)}$ are supposed to exist, satisfying the equation

$$(4.4.2) \quad \int_a^b f(t) dy^{(N-1)}(t) - a_1 \int_a^b f(t) dy^{(N-2)}(t) - \dots \\ - a_N \int_a^b f(t) y(t) dt = c \int_a^b f(t) d\xi(t)$$

with probability 1, for each continuous function $f(t)$ and each pair of numbers a, b . The formal integrals are defined as the limit in the mean of the usual sums.²⁰ The integral on the right has already been discussed. With this interpretation, equations involving ξ' can be treated in the usual way, and this will be done in the following without further comment. The formal solution of (4.4.2) is well known. Let $\lambda_1, \dots, \lambda_N$ be the roots of the equation

$$(4.4.3) \quad \lambda^N - a_1 \lambda^{N-1} - \dots - a_N = 0$$

and suppose that these roots are distinct, and have negative real parts. Let $\Delta_{j,k}$ be the cofactor of λ_j^{k-1} in the determinant

$$(4.4.4) \quad \delta = \begin{vmatrix} 1 & \cdot & \cdot & \cdot & 1 \\ \lambda_1 & \lambda_2 & \cdot & \cdot & \lambda_N \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \lambda_1^{N-1} & \lambda_2^{N-1} & \cdot & \cdot & \lambda_N^{N-1} \end{vmatrix}.$$

Then the general solution of (4.4.1), that is to say of (4.4.2), is

$$(4.4.5) \quad y(t) = \frac{c}{\delta} \int_0^t \sum_{j=1}^N \Delta_{N,j} e^{\lambda_j(t-s)} d\xi(s) + \frac{1}{\delta} \sum_{j,k=1}^N \Delta_{k,j} e^{\lambda_j t} y^{(k-1)}(0).$$

Since the integrand and its first $N-1$ derivatives vanish when $s = t$, $y', \dots, y^{(N-1)}$ as defined by (4.4.5) exist, but $y^{(N)}(t)$ does not exist, because $\xi'(t)$ in (4.4.1)

²⁰ For a full discussion in the case $N = 1$ cf. Doob, *Annals of Math.*, Vol. 43 (1942), pp. 358-61.

does not exist. The $y(t)$ process is a t.h.G. process if $y(0), \dots, y^{(N-1)}(0)$ are chosen properly. This can be seen from the solution

$$(4.4.6) \quad y(t) = \frac{c}{\delta} \int_{-\infty}^t \sum_{j=1}^N \Delta_{Nj} e^{\lambda_j(t-s)} d\xi(s).$$

In fact this is the only solution defining a t.h.G. process. To prove this, rewrite (4.4.5) in the form

$$(4.4.5') \quad y(t) = \frac{c}{\delta} \sum_{j=1}^N \Delta_{Nj} \int_{1-\tau}^t e^{\lambda_j(t-s)} d\xi(s) + \frac{1}{\delta} \sum_{j,k=1}^N \Delta_{kj} e^{\lambda_j(t-\tau)} y^{(k-1)}(\tau).$$

If the $y(t)$ process is a t.h.G. process, (4.4.5') becomes (4.4.6) when $\tau \rightarrow -\infty$. Thus there is a unique stationary solution to (4.4.1) and, by (4.4.5), every solution tends to this solution in the long run. The stationary solution (4.4.6) has the property that $y(t)$ is written in terms of $\xi(s)$ for $s \leq t$. Then in (4.4.5) the integral is independent of the terms involving the initial conditions. In other words

$$(4.4.7) \quad E\{y(s), s \leq 0; y(t)\} = \frac{1}{\delta} \sum_{j,k=1}^N \Delta_{kj} e^{\lambda_j t} y^{(k-1)}(0).$$

Hence the variables $y(t), y'(t), \dots, y^{(N-1)}(t)$ define an N -dimensional t.h.G.M. process. The transition matrix function $A(t)$, and the matrices Q, S, T of Theorem 4.3 are easily calculated.

$$(4.4.8) \quad \begin{aligned} A(t) &: (\Delta_{kj} \lambda_j^{k-1} e^{\lambda_j t}), \\ Q &: \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 \\ a_N & \cdot & \cdot & \cdot & a_1 \end{pmatrix} \\ T &= 0. \end{aligned}$$

The necessary changes to be made if the λ_j are not distinct are well known. The case $c = 0$ will be treated below, when the problem will be reconsidered from another point of view. In all cases the solution of (4.4.1) leads to an N -dimensional t.h.G.M. process.²¹

As a simple example, consider a torsion pendulum, suspended in a sealed container. The only turning forces acting on the pendulum are the molecular shocks of the surrounding gas, and the restoring torque. The equation of motion is

$$(4.4.9) \quad I \frac{d^2 y(t)}{dt^2} + \alpha_1 \frac{dy(t)}{dt} + \alpha_2 y(t) = X(t),$$

²¹ According to a letter from Uhlenbeck, the differential equation (4.4.1) was solved; from a somewhat different point of view, by Miss Ming Chen Wang, in a thesis written in 1941 which is unfortunately inaccessible to me at the moment.

where y is the angular displacement measured from the equilibrium position, I is the moment of inertia, α_2 is the torque coefficient of the suspension, and the molecular force is resolved into a systematic Stokes term $\alpha_1 y'$ and a remainder X . The remainder term $X(t)$ defines a stationary process which to a first approximation is "purely random." In the present context "purely random" means that if $t_1 < \dots < t_n$, $X(t_1), \dots, X(t_n)$ are mutually independent. This is precisely the property the derived process of a ξ -process would have, if $\xi'(t)$ existed. Unfortunately it has already been noted that $\xi'(t)$ does not exist, since the difference quotient $[\xi(t+h) - \xi(t)]/h$ is unbounded as $h \rightarrow 0$. It has already been seen, however, that (4.4.9) can be given a meaning with $X(t)$ identified with $c\xi'(t)$ even though $\xi'(t)$ does not exist, and it has been seen that the solution approaches a steady state. It may still be a disappointment to some that the solution $y(t)$ has a first derivative $y'(t)$ but that $y''(t)$ does not exist: there is an angular velocity but not an angular acceleration! This unhappy circumstance can either be blamed on the physical world, or on the mathematical approximation to the physical world, depending on the point of view. The corresponding electrical picture is the following. There are spontaneous currents in any electrical circuit, due to the thermal motion of the electrons. This is known as the Johnson effect. In a simple closed circuit, consisting of an inductance L , a resistance, R , and a capacitance C in series, the current equation can be written in the form

$$(4.4.10) \quad L \frac{d^2 y(t)}{dt^2} + R \frac{dy(t)}{dt} + \frac{y(t)}{C} = X(t),$$

where y is the charge on the condenser and $X(t)$ represents a fictitious voltage set up by the motion of the electrons. The $X(t)$ is identified with $c\xi'(t)$ as before. In this case there is a current $\frac{dy}{dt}$, but the current function has no derivative. In these applications, the physical justification for the Gaussian character of the ξ -distribution lies in the Gaussian character of the Maxwell distribution of elementary particle velocities. The known mean particle kinetic energy determines the constant c in (4.4.1). The more complicated mechanical or electrical systems will lead to equations of higher order than 2, or systems of equations. For example the usual current equations of a net or resistances capacitances and inductances lead to a system of say ν second order equations of type (4.4.10), and the corresponding pairs y, y' form a 2ν -dimensional t.h.G.M. process.²²

The processes defined by linear differential equations of the type (4.4.1) are the analogues of the t.h.G.M._N. processes in the discrete parameter case. Instead of defining these solutions of (4.4.1) as the t.h.G.M._N. processes, however, we shall use a definition closer to the definition in the discrete parameter case. A

²² Further discussion and references to papers by physicists on this subject will be found in Doob, *Annals of Math.*, Vol. 43 (1942), pp. 351-69.

one-dimensional t.h.G. process with variables $\{y(t)\}$ will be called a t.h.G.M._N process if the derivatives $y'(t), \dots, y^{(N-1)}(t)$ exist, and if whenever $s < t$,

$$(4.5.1) \quad E\{y(\tau), \tau \leq s; y(t)\} = E\{y(s), y'(s), \dots, y^{(N-1)}(s); y(t)\}.$$

If $N = 1$, the process is a t.h.G.M. process. The right hand side of (4.5.1) is a linear combination of the variables $y(s), \dots, y^{(N-1)}(s)$. The variables $\{y(t)\}$ thus satisfy an equation of the form

$$(4.5.2) \quad y(t) - a_1(t-s)y(s) - \dots - a_N(t-s)y^{(N-1)}(s) = \eta(s, t)$$

where $\eta(s, t)$ is independent of the variables $\{y(\tau)\}$ for $\tau \leq s$. Define the variables $\{x(t)\}$ of an N -dimensional t.h.G. process by

$$(4.5.3) \quad \begin{aligned} x_1(t) &= y(t) \\ x_j(t) &= y^{(j)}(t), \quad j = 1, \dots, N-1. \end{aligned}$$

If this process is degenerate, there is a relation of the form

$$(4.5.4) \quad c_0 y(s) + c_1 y'(s) + \dots + c_{N-1} y^{(N-1)}(s) = 0, \quad \sum_{j=0}^{N-1} |c_j| > 0.$$

It can be assumed that $c_{N-1} \neq 0$, (differentiating (4.5.4) to get a term in $y^{(N-1)}(t)$ if there is none originally). Then $y^{(N-1)}(s)$ can be eliminated in (4.5.2), to get a relation of the same type with N replaced by $N-1$. Hence the process is non-degenerate if N is the minimum index for which (4.5.1) is true. It will now be proved that the $x(t)$ process is a t.h.G.M. process. It can be assumed to be non-degenerate. Using (4.5.1),

$$(4.5.5) \quad E\{x(\tau), \tau \leq s; x_1(t)\} = E\{y(\tau), \tau \leq s; y(t)\} = E\{x(s); x_1(t)\}.$$

It must also be shown that

$$(4.5.6) \quad E\{x(\tau), \tau \leq s; x_j(t)\} = E\{x(s); x_j(t)\} \quad j = 2, \dots, N.$$

This will be shown by justifying the taking of derivatives in (4.5.5). It will be sufficient to prove (4.5.6) when $j = 2$. Using (4.5.1),

$$(4.5.7) \quad E\left\{x(\tau), \tau \leq s, \frac{y(t+h) - y(t)}{h}\right\} = E\left\{x(s); \frac{y(t+h) - y(t)}{h}\right\}.$$

The right hand side is a linear combination of $x_1(s), \dots, x_N(s)$ whose coefficients are continuous in h , $h \geq 0$, since the correlation function of the $y(t)$ process is continuous. Hence the right hand side converges to

$$E\{x(s); y'(t)\} = E\{x(s); x_2(t)\}$$

when $h \rightarrow 0$. Since the difference

$$x(t) - E\left\{x(s); \frac{y(t+h) - y(t)}{h}\right\}$$

is uncorrelated with $x(\tau)$ if $\tau \leq s$, the same is true of its limit as $h \rightarrow 0$. This means that (4.5.6) is true when $j = 2$, as was to be shown. Conversely if $\{y(t)\}$ are the variables of a one-dimensional t.h.G. process, if $y'(t), \dots, y^{(N-1)}(t)$ exist, and if the $x(t)$ process defined by (4.5.3) is a t.h.G.M. process, the $y(t)$ process is obviously a t.h.G.M._N process. The transition matrix function $A(t)$ and the matrices Q, S, T of Theorem 4.3 are easily calculated. Suppose that the $x(t)$ process is non-degenerate. Since $y^{(i-1)}(t)$ is given by

$$\begin{aligned} y^{(i-1)}(t) &= x_i(t) = \int_{-\infty}^t \sum_{j=1}^N [e^{(t-s)Q} S]_{ij} d\xi_j(s) + \sum_{j=1}^N [e^{tQ} T]_{ij} \xi_j \\ (4.5.8) \quad &= \int_0^t \sum_{j=1}^N [e^{(t-s)Q} S]_{ij} d\xi_j(s) + \sum_{j=1}^N (e^{tQ})_{ij} x_j(0) \end{aligned}$$

and since $x'_i(t)$ exists if $i < N$, it follows that the integrand must vanish when $s = t$:

$$(4.5.9) \quad (S)_{ii} = 0, \quad i = 1, \dots, N-1, \quad j = 1, \dots, N.$$

Since S is symmetric and non-negative definite, S must have the form

$$(4.5.10) \quad S: \begin{pmatrix} 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & c \end{pmatrix}, \quad c \geq 0.$$

The fact that $x'_i(t) = x_{i+1}(t)$ means that

$$(4.5.11) \quad \sum_{j=1}^N (e^{tQ} Q)_{ij} x_j(0) = \sum_{j=1}^N (e^{tQ})_{i+1j} x_j(0), \quad j = 1, \dots, N$$

or, since the $x(t)$ process is non-degenerate,

$$(4.5.12) \quad (e^{tQ} Q)_{ij} = (e^{tQ})_{i+1j} \quad \begin{matrix} i = 1, \dots, N-1 \\ j = 1, \dots, N. \end{matrix}$$

Hence $(t \rightarrow 0)$ Q has the form

$$(4.5.13) \quad Q: \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 \\ a_N & \cdot & \cdot & \cdot & \cdot & a_1 \end{pmatrix}.$$

Conversely if there is an N -dimensional non-degenerate t.h.G.M. process with transition matrix function e^{tQ} where Q is given by (4.5.13) and dispersion matrix S given by (4.5.10),

$$x'_i(t) = x_{i+1}(t), \quad i = 1, \dots, N-1,$$

and the $x_1(t)$ process is a t.h.G.M._N process.

Case 1. $S = 0$ (deterministic case). In this case the $x(t)$ process is deterministic:

$$(4.5.14) \quad x(t) = e^{tQ} T \xi.$$

Since Q satisfies its characteristic equation

$$(4.5.15) \quad \alpha^N - a_1 \alpha^{N-1} - \dots - a_N = 0$$

it follows that

$$(4.5.16) \quad x^{(N)}(t) - a_1 x^{(N-1)}(t) - \dots - a_N x(t) = 0,$$

$$(4.5.17) \quad y^{(N)}(t) - a_1 y^{(N-1)}(t) - \dots - a_N y(t) = 0.$$

The roots of (4.5.15) are simple roots, and are all pure imaginary, according to Theorem 4.3. It follows that

$$(4.5.18) \quad y(t) = \sum_j (\eta_j \cos t\theta_j + \zeta_j \sin t\theta_j)$$

where the η_j and ζ_j are one-dimensional Gaussian variables, and $\{i\theta_j\}$ are the distinct roots of (4.5.15).

Case 2. $S \neq 0$ (non-deterministic case). In this case it will now be shown that the $x(t)$ process has no deterministic factor, that is that the roots of (4.5.15) all have negative real parts. In fact let β be a root of (4.5.15), corresponding to the characteristic vector z of Q^* :

$$(4.5.19) \quad z = (a_N \beta^{N-1}, a_N \beta^{N-2} + a_{N-1} \beta^{N-1}, \dots, a_N + a_{N-1} \beta + \dots + a_1 \beta^{N-1}) \\ = (\dots, \beta^N).$$

Then using (4.3.5)

$$(4.5.20) \quad 0 < c^2 |\beta|^{2N} = (S^2 z, z) = -(R(0)Q^* z, z) - (QR(0)z, z) \\ = -\lambda(R(0)z, z) - \bar{\lambda}(R(0)z, z) \\ = -(\lambda + \bar{\lambda})(R(0)z, z).$$

Hence $\lambda + \bar{\lambda}$ is real and negative: λ has a negative real part. In this non-deterministic case, therefore, the $x(t)$ process can have no deterministic factor. The matrix T is the null matrix, and (4.3.2) becomes

$$(4.5.21) \quad x(t) = \int_{-\infty}^t e^{(t-s)Q} S d\xi(s)$$

which leads to

$$(4.5.22) \quad y(t) = c \int_{-\infty}^t [e^{(t-s)Q}]_{1N} d\xi_N(s).$$

is a ξ -process, if c is chosen properly, unless the parenthesis in (4.6.9) vanishes for all j . In either case (iii) is proved.

Proof of (iv). If in (iii), $x_r^{(N)}(t)$ exists, (4.6.3) can be augmented to include

$$(4.6.3') \quad r_r^{(N)}(t) = \int_{-\infty}^t \sum_{j=1}^N [e^{(t-s)Q} Q^N S]_{rj} d\xi_j(s) + \sum_{j=1}^N [e^{tQ} Q^N] x_j(0)$$

and (4.6.4) now includes

$$(4.6.4') \quad (Q^{N-1}S)_{rj} = 0, \quad j = 1, \dots, N.$$

In this case the last term in (4.6.7) vanishes and (4.6.8), with zero on the right hand side, is strictly true

Proof of (v). If $x_1'(t), \dots, x_N'(t)$ exist, S must vanish and (4.6.3) yields

$$(4.6.10) \quad x(t) = e^{tQ} x(0), \quad x^{(v)}(t) = Q^v x(t)$$

Thus the $x(t)$ process is deterministic and $x(t)$ has derivatives of all orders

THEOREM 4.7. *Let $\{x(t)\}$ be the variables of a one-dimensional t.h.G. process. The process is a component process of an N -dimensional t.h.G.M. process if and only if the chance variables*

$$(4.7.1) \quad x(0), \{E\{x(s), s \leq 0; x(t)\}\} \quad 0 < t < \infty$$

are linearly dependent on N variables.

Suppose that the $x(t)$ process is a component process of an N -dimensional t.h.G.M. $y(t)$ process: $x(t) = y_r(t)$, and let $A(t)$ be the transition matrix function of the $y(t)$ process. Then if $\epsilon > 0$ and if n is any integer, the difference

$$y[(n+1)\epsilon] - A(\epsilon)y(n\epsilon)$$

is independent of every $y(s)$ with $s \leq n\epsilon$, and therefore independent of every $y(m\epsilon)$ with $m \leq n$. Hence the $y(n\epsilon)$ process is a t.h.G.M. process (discrete parameter case). Equation (3.8.5) becomes, in this case, if $n = 0$,

$$(4.7.2) \quad x[(N+\nu)\epsilon] - a_1^{(\nu)} x[(N-1)\epsilon] - \dots - a_N^{(\nu)} x(0) = \eta_1^{(\nu)}(N+\nu)$$

where $\eta_1^{(\nu)}(N+\nu)$ is not merely independent of the variables $\dots, x(-\epsilon), x(0)$, but is even independent of every $x(s)$ with $s \leq 0$. It then follows, applying the operator $E\{x(s), s \leq 0; \cdot\}$ to both sides of (4.7.2), that the variables in (4.7.1) are linearly dependent on N variables if t is restricted to be a multiple of ϵ . Allowing ϵ to run through the values

$$\left\{ \frac{1}{m!} \right\}, \quad m = 1, 2, \dots$$

it follows that the statement of the theorem is true if t is restricted to be rational. The proof will be complete when it is shown that the subject²³ variables for rational t are dense in the whole class in the sense that for any t , the expectation

$$(4.7.3) \quad \delta = E\{[E\{x(s), s \leq 0; x(t')\} - E\{x(s), s \leq 0; x(t)\}]^2\}$$

²³ Courtesy of U. S. Navy

converges to 0 when $t' \rightarrow t$. In fact, using the Schwarz inequality

$$(4.7.4) \quad \begin{aligned} \delta &= E\{[E\{x(s), s \leq 0; x(t') - x(t)\}]^2\} \\ &\leq E\{E\{x(s), s \leq 0, [x(t') - x(t)]^2\}\} = E\{[x(t') - x(t)]^2\} \end{aligned}$$

and the basic continuity hypothesis (1.3.1) imposed on continuous processes is precisely that the last expectation converges to 0 when $t' \rightarrow t$.

Conversely suppose that the chance variables (4.7.1) are linearly dependent on N variables. It can be supposed that $x(0)$ is one of these N . Let the others be those for which $t = t_2, \dots, t_N$, and define $y_1(t), \dots, y_N(t)$ by

$$(4.7.5) \quad \begin{aligned} y_1(t) &= x(t) \\ y_j(t) &= E\{x(s), s \leq t; x(t + t_j)\} \quad j = 2, \dots, N. \end{aligned}$$

The $y(t)$ process is obviously an N -dimensional t.h.G. process. Moreover

$$(4.7.6) \quad \begin{aligned} E\{y(s), s \leq 0; y_j(t) = E\{x(s), s \leq 0; y_j(t)\} \\ = E\{x(s), s \leq 0; x(t + t_j)\} \quad j = 1, \dots, N \end{aligned}$$

(where t_1 is defined as 0). Since the right side is by hypothesis, for each j , a linear combination of $y_1(0), \dots, y_N(0)$, the $y(t)$ process is a t.h.G.M. process, and the $x(t)$ process is a component process, as was to be shown.

A detailed examination will now be made of t.h.G.M. processes, and of the more general class of component processes of t.h.G.M. processes. The following theorem will be useful.

THEOREM 4.8. *Let $\{x(t)\}$ be the variables determining a t.h.G. continuous parameter process. The process is a component process of an N -dimensional t.h.G.M. process if and only if for each $\epsilon > 0$ the discrete parameter process with variables $\{x(n\epsilon)\}$ is a component process of an N -dimensional t.h.G.M. process.*

If the $x(t)$ process is a component process of an N -dimensional t.h.G.M. $y(t)$ process, the $x(n\epsilon)$ process is a component process of the N -dimensional t.h.G.M. $y(n\epsilon)$ process. Conversely suppose that the $x(n\epsilon)$ process is a component process of an N -dimensional t.h.G.M. process (which may depend on ϵ) for every $\epsilon > 0$. It follows that for each $\epsilon > 0$ the chance variables

$$(4.8.1) \quad E\{\dots, x(-\epsilon), x(0); x(n\epsilon)\}, \quad n = 0, 1, \dots$$

are linearly dependent on N of their number. Hence the same is true of the following chance variables, if ν, m are fixed and $\nu > m$:

$$(4.8.2) \quad E\{\dots, x(-1/\nu), x(0); x(n/m)\} \quad n = 0, 1, \dots$$

According to the Corollary to Theorem 1.2, when $\nu \rightarrow \infty$ the conditional expectations in (4.8.2) converge to

$$(4.8.3) \quad E\{x(s), s \leq 0, s \text{ rational}; x(n/m)\} \quad n = 0, 1, \dots$$

Hence the chance variables (t rational)

$$(4.8.4) \quad E\{x(s), s \leq 0, s \text{ rational}; x(t)\} = E\{x(s), s \leq 0; x(t)\}^{24}, \quad 0 < t < \infty,$$

are linearly dependent on N of their number. As in the proof of Theorem 4.7 it follows that the same is true if t runs through all positive real numbers, and according to Theorem 4.7, the $x(t)$ process is therefore a component process of an N -dimensional t.h.G.M. process.

THEOREM 4.9. *Let $\{x(t)\}$ be the variables of a one-dimensional continuous parameter t.h.G. process. The process is a component process of a finite-dimensional t.h.G.M. process if and only if the complex spectral function of the process is the sum of the integral of the square of the absolute value of a rational function of λ and of a monotone non-decreasing function increasing only in a finite number of jumps.²⁵ Specifically:*

(i) *The process is a component process of an N -dimensional t.h.G.M. process if and only if the complex spectral function has the form*

$$(4.9.1) \quad G(\lambda) = \int_{-\infty}^{\lambda} \frac{|\beta_0(i\lambda)^{N-1} + \dots + \beta_{N-1}|^2}{|(i\lambda)^N + \alpha_1(i\lambda)^{N-1} + \dots + \alpha_N|^2} d\lambda + \hat{G}(\lambda)$$

where

(a) $\hat{G}(\lambda)$ is a monotone non-decreasing function satisfying (1.3.3) and increasing only in jumps, at no more than N points.

(b) the denominator of the integrand vanishes at every discontinuity of $\hat{G}(\lambda)$, and the numerator vanishes at every zero of the denominator, to at least the same order;

(c) the coefficients in the integrand are real, and the roots of the λ polynomials are all on the real axis or in the upper half plane

The integral vanishes identically if and only if the $x(n)$ process is a component process of an N -dimensional deterministic process, and $\hat{G}(\lambda)$ vanishes identically if and only if the variables $\{x(t)\}$ vanish identically or the $x(t)$ process is a component process of an N -dimensional t.h.G.M. process with no deterministic factor.

(ii) *The process is a t.h.G.M._N process, in the deterministic case, if and only if the complex spectral function $G(\lambda) = \hat{G}(\lambda)$ is a function increasing only in jumps, at no more than N points; non-deterministic case if and only if the complex spectral function has the form*

$$(4.9.2) \quad G(\lambda) = \int_{-\infty}^{\lambda} \frac{c d\lambda}{|(i\lambda)^N + \dots + \alpha_N|^2}.$$

²⁴ The equality (4.8.4) is proved as follows. Let t be fixed, and let x be the chance variable on the left. Then $x(t) - x$ has mean 0 and is uncorrelated with every $x(s)$ with $s \leq 0$ and rational. It follows at once from the continuity of hypothesis (1.3.1) that then $x(t) - x$ is uncorrelated with every $x(s)$ with $s \leq 0$. It follows that (4.8.4) is true.

²⁵ It is easily seen that the first term of the two can also be described simply as the integral of a rational function of λ , which is non-negative for real λ and is integrable and an even function, like all complex spectral density functions.

Proof of (i). Suppose that the $x(t)$ process is a one-dimensional component process of an N -dimensional t.h.G.M. $y(t)$ process, $x(t) = y_1(t)$. It is no restriction to assume that the $y(t)$ process is non-singular. Then the correlation function of the $y(t)$ process is given by

$$(4.9.3) \quad \begin{aligned} R_y(t) &= R_y(0)e^{iQ^*t} & t \geq 0 \\ R_y(t) &= e^{-iQ} R_y(0) & t \leq 0, \end{aligned}$$

where Q is uniquely determined and

$$(4.9.4) \quad \begin{aligned} G(\lambda) - G(0) &= \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i\lambda t} - 1}{it} R_y(t) dt \right]_{11} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i\lambda t} - 1}{it} [R_y(t)]_{11} dt, \end{aligned}$$

at the points of continuity of $G(\lambda)$.

The correlation function $R_y(t)$ has derivatives of all orders for $t > 0$:

$$(4.9.5) \quad \begin{aligned} R_y^{(r)}(t) &= R_y(0)Q^{*r}e^{iQ^*t} & t > 0 \\ &= (-1)^r Q^r e^{-iQ} R_y(0) & t < 0. \end{aligned}$$

Suppose first that the $y(t)$ process has no deterministic factor, in other words that it is non-degenerate and of type M . Then the characteristic values of Q have negative real parts and $R(t) \rightarrow 0$ exponentially when $|t| \rightarrow \infty$. Hence $G(\lambda)$ has a continuous derivative $G'(\lambda)$:

$$(4.9.6) \quad G'(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\lambda t} [R_y(t)]_{11} dt.$$

Integrating by parts,

$$(4.9.7) \quad \begin{aligned} G'(\lambda) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i\lambda t}}{i\lambda} [R_y'(t)]_{11} dt \\ &= \frac{R_y'(0+) - R_y'(0-)}{2\pi(i\lambda)^2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i\lambda t}}{(i\lambda)^2} R_y''(t) dt \\ &= \frac{R_y'(0+) - R_y'(0-)}{2\pi(i\lambda)^2} - \frac{R_y''(0+) - R_y''(0-)}{2\pi(i\lambda)^3} - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i\lambda t}}{(i\lambda)^3} R_y'''(t) dt, \end{aligned}$$

Since Q satisfies its characteristic equation

$$(4.9.8) \quad \alpha^N - a_1 \alpha^{N-1} - \dots - a_N = 0,$$

it follows that

$$(4.9.9) \quad \begin{aligned} R_y^{(N)}(t) - a_1 R_y^{(N-1)}(t) - \dots - a_N R_y(t) &= 0 & t > 0 \\ R_y^{(N)}(t) + a_1 R_y^{(N-1)}(t) - \dots + (-1)^{N-1} R_y(t) &= 0 & t < 0 \end{aligned}$$

and therefore if U is the operator $\frac{d}{dt}$

$$(4.9.10) \quad [U^N - a_1 U^{N-1} - \dots - a_N U^0][U^N + a_1 U^{N-1} - \dots + (-1)^{N-1} U^0] R_y(t) = 0, \quad t \neq 0.$$

Applying (4.9.10) to (4.9.6)

$$(4.9.11) \quad \begin{aligned} & [(i\lambda)^N - a_1(i\lambda)^{N-1} - \dots - a_N][(i\lambda)^N + a_1(i\lambda)^{N-1} + \dots \\ & \quad + (-1)^{N-1} a_N] G'(\lambda) \\ & = |(i\lambda)^N - a_1(i\lambda)^{N-1} - \dots - a_N|^2 G'(\lambda) = P(i\lambda) \end{aligned}$$

where $P(i\lambda)$ is a polynomial of degree $2n - 2$. Since $P(i\lambda)$ is real and non-negative, when λ is real, the roots on the real axis are of even multiplicity and those off the axis are symmetric in the axis. Moreover $P(i\lambda)$ is even, since the left side of (4.9.11) is even. It follows easily that $P(i\lambda)$ can be written in the form

$$(4.9.12) \quad P(i\lambda) = \left| \sum_{j=0}^{N-1} \beta_j (i\lambda)^{N-j-1} \right|^2$$

where the roots of the β polynomial are all on or to the left of the imaginary axis. Finally

$$(4.9.13) \quad G'(\lambda) = \frac{|\beta_0(i\lambda)^{N-1} + \dots + \beta_{N-1}|^2}{|(i\lambda)^N - \dots - a_N|^2}.$$

The denominator polynomial in λ vanishes only at points where $i\lambda$ has a negative real part, that is where λ has a positive imaginary part. This completes the proof in the case where the N -dimensional $y(t)$ process has no deterministic factor. If there are such factors, it is easily verified that $G(\lambda)$ has corresponding discontinuities and the above proof then applies to $G(\lambda)$ less its jump function. The result can finally be summarized as in the statement of the theorem. If the $y(t)$ process has only deterministic factors $[R_y(t)]_{11}$ will be a sum of trigonometric functions and $G(\lambda)$ will be a function of jumps.

Conversely suppose that the $x(t)$ process has the complex spectral function (4.9.13). Then following the ideas of the proof of the analogous section of Theorem 3.9, it follows that $R(t)$ satisfies the differential equation (cf. (3.9.10) and (3.9.11)).

$$(4.9.14) \quad R^{(N)}(t) - a_1 R^{(N-1)}(t) - \dots - a_N R(t) = 0, \quad t > 0.$$

Any solution of (4.9.14) is a linear combination of (at most N) functions

$$(4.9.15) \quad e^{\beta t}, t e^{\beta t}, \dots$$

where β is a root of the equation

$$(4.9.16) \quad \alpha^N - a_1 \alpha^{N-1} - \dots - a_N = 0$$

and where powers of t may appear if β is a multiple root. Let ϵ be a positive number. The discrete parameter process determined by the variables $\{x(n\epsilon)\}$

has correlation function $R(n\epsilon)$. This function is a linear combination of function

$$(4.9.15') \quad (e^{\beta_1 \epsilon})^n, n(e^{\beta_2 \epsilon})^n, \dots$$

corresponding to those of (4.9.15). There is an equation

$$(4.9.17) \quad \alpha^N - a_1(\epsilon)\alpha^{N-1} - \dots - a_N(\epsilon) = 0$$

with the $\{e^{\beta_i \epsilon}\}$ as roots, of the same multiplicity as that of β in (4.9.16). Hence

$$(4.9.18) \quad R[(n+N)\epsilon] - a_1(\epsilon)R[(n+N-1)\epsilon] - \dots - a_N(\epsilon)R(n\epsilon) = 0, \\ n \geq 0.$$

According to Theorem 3.10 the $x(n\epsilon)$ discrete parameter process is therefore a component process of an N -dimensional discrete parameter t.h.G.M. process. Since this is true for all ϵ , the $x(t)$ process is a component process of an N -dimensional continuous parameter t.h.G.M. process.

If the integral vanishes identically, the non-deterministic factors in the N -dimensional process are irrelevant to the $x(t)$ process and can be replaced by factors of type $M(0)$. If on the other hand the spectral function is continuous, the deterministic factors are irrelevant and can be replaced by factors of type M .

Proof of (ii). Since the t.h.G.M. $_N$ processes are characterized among the component processes of N -dimensional t.h.G.M. processes by the fact that the first $N-1$ derived process exist, their spectral functions (according to Theorem 1.4) are characterized by the property that

$$\int_{-\infty}^{\infty} \lambda^{2(N-1)} dG(\lambda) < \infty$$

that is the numerator in (4.9.1) must be identically constant. If this constant is not 0, $\hat{G}(\lambda)$ can have no jumps, since each jump corresponds to a zero of numerator and denominator. Hence $G(\lambda)$ is either identically $\hat{G}(\lambda)$ or is in the form (4.9.2). The two possibilities obviously correspond to the deterministic and non-deterministic cases, respectively.

COROLLARY. *The t.h.G.M. $_N$ one dimensional process which is the solution of (4.4.1) has complex spectral function*

$$(4.9.19) \quad \int_{-\infty}^{\lambda} \frac{c^2 d\lambda}{|(i\lambda)^N - a_1(i\lambda)^{N-1} - \dots - a_N|^2}.$$

In fact the complex spectral function has the form (4.9.2), where the coefficients in the polynomial are those of the differential equation for the correlation function $R_\epsilon(t)$ in (4.9.9), that is the coefficients of the characteristic equation of the infinitesimal transition matrix Q , (cf. (4.4.8)). The evaluation (4.9.19) is also easily proved directly.

The analogues of Theorems 3.10 and 3.11 in the continuous parameter case are easy to prove and will be omitted.

ON CUMULATIVE SUMS OF RANDOM VARIABLES

BY ABRAHAM WALD

Columbia University

1. Introduction. Let $\{z_i\}$ ($i = 1, 2, \dots$, ad inf.) be a sequence of independent random variables each having the same distribution. Denote by Z_j the sum of the first j elements of the sequence $\{Z_i\}$, i.e.,

$$(1) \quad Z_j = z_1 + z_2 + \dots + z_j, \quad (j = 1, 2, \dots, \text{ad inf.}).$$

Let a be a given positive constant and b a given negative constant. Denote by n the smallest positive integer for which Z_n lies outside the open interval (b, a) , i.e., Z_n is either $\leq b$ or $\geq a$. Obviously n is a random variable. If $b < Z_i < a$ for $i = 1, 2, \dots$, ad inf., we shall say that $n = \infty$.

For any relation R we shall denote the probability that R holds by $P(R)$. It will be shown later that $P(n = \infty) = 0$, provided the variance of z_i is positive.

In this paper we shall deal with the problem of obtaining the value of $P(Z_n \geq a)$ ¹ and that of finding the probability distribution of n .

The study of such cumulative sums is of interest in various statistical problems. For example, a multiple sampling scheme proposed recently by Walter Bartky² makes use of such cumulative sums.

Cumulative sums also play an important role in the theory of the random walk of interest in physics. The results obtained in this paper may have bearing particularly on the theory of the random walk with absorbing barriers. In the presence of an absorbing wall the random walk stops whenever the particle arrives at the wall, i.e., whenever the cumulative sum of the displacements reaches a certain value.³

2. Two Lemmas. LEMMA 1. *If the variance of z_i is not zero, $P(n = \infty) = 0$.*

PROOF: Let $c = |a| + |b|$. If $n = \infty$ then for any positive integer r the following inequalities must hold

$$(2) \quad \left(\sum_{i=kr+1}^{(k+1)r} z_i \right)^2 < c^2 \quad (k = 0, 1, 2, \dots, \text{ad inf.}).$$

To prove $P(n = \infty) = 0$, it is sufficient to show that the probability is zero that (2) holds for all integer values of k . Since the variance of z_i is not zero, the ex-

¹ Since $P(n = \infty) = 0$, we have $P(Z_n \leq b) = 1 - P(Z_n \geq a)$

² "Multiple sampling with constant probability", *Annals of Math. Stat.*, Vol. 14 (1943), pp. 363-377.

³ See in this connection S. Chandrasekhar, "Stochastic problems in physics and astronomy", *Rev. of Modern Physics*, Vol. 15 (1943), p. 5

pected value of $\left(\sum_{i=1}^j z_i\right)^2$ converges to ∞ as $j \rightarrow \infty$. Hence there exists a positive integer r such that

$$(3) \quad P\left[\left(\sum_{i=1}^r z_i\right)^2 < c^2\right] < 1.$$

From (3) it follows that the probability that (2) is fulfilled for all values of k is equal to zero. Hence $P(n = \infty) = 0$ and Lemma 1 is proved.

LEMMA 2. Let z be a random variable such that the following four conditions are fulfilled:

Condition I. Both the expected value Ez of z and the variance of z exist and are unequal to zero.

Condition II. There exists a positive δ such that $P(e^z < 1 - \delta) > 0$ and $P(e^z > 1 + \delta) > 0$.

Condition III. For any real value h the expected value $Ee^{hz} = g(h)$ exists.

Condition IV. The first two derivatives of the function $g(h)$ exist and may be obtained by differentiation under the integral sign, i.e.,

$$g'(h) = \frac{d}{dh} Ee^{hz} = Eze^{hz},$$

and

$$g''(h) = \frac{d^2}{dh^2} Ee^{hz} = Ez^2 e^{hz}.$$

Then there exists one and only one real value $h_0 \neq 0$ such that

$$Ee^{h_0 z} = 1.$$

PROOF: For any positive h we have

$$(4) \quad g(h) > P(e^z > 1 + \delta)(1 + \delta)^h.$$

Hence, since $P(e^z > 1 + \delta) > 0$,

$$(5) \quad \lim_{h \rightarrow \infty} g(h) = +\infty.$$

Similarly we see that for any negative h

$$(6) \quad g(h) > P(e^z < 1 - \delta)(1 - \delta)^h.$$

Hence, since $P(e^z < 1 - \delta) > 0$, we have

$$(7) \quad \lim_{h \rightarrow -\infty} g(h) = +\infty.$$

Since $g''(h) = Ez^2 e^{hz}$ it follows easily from Condition II that

$$(8) \quad g''(h) > 0,$$

for all real values of h .

The relations (5), (7) and (8) imply that there exists exactly one real value h^* for which $g(h)$ takes its minimum value. Since $g'(0) = Ez$ is unequal to zero by Condition I, we see that $h^* \neq 0$ and $g(h^*) < g(0) = 1$. It is clear that the function $g(h)$ is monotonically decreasing in the strict sense over the interval $(-\infty, h^*)$, and is monotonically increasing in the strict sense over the interval $(h^*, +\infty)$. Since $g(0) = 1$ and $g(h^*) < 1$, there exists exactly one real value $h_0 \neq 0$ such that $g(h_0) = 1$. Hence Lemma 2 is proved.

3. A fundamental identity. Denote by z a random variable whose distribution is equal to the common distribution of $z_i (i = 1, 2, \dots, \text{ad inf.})$. Let D' be the subset of the complex plane such that $Ee^{zt} = \varphi(t)$ exists and is finite for any point t in D' . Consider the following identity

$$(9) \quad Ee^{zn^t + (z_N - z_n)t} = Ee^{z_N t} = [\varphi(t)]^N,$$

where N denotes a positive integer. Let P_N be the probability that $n \leq N$. For any random variable u denote by $E_N(u)$ the conditional expected value of u under the restriction that $n \leq N$, and by $E_N^*(u)$ the conditional expected value of u under the restriction that $n > N$. Then identity (9) can be written as

$$(10) \quad P_N E_N e^{zn^t + (z_N - z_n)t} + (1 - P_N) E_N^* e^{z_N t} = [\varphi(t)]^N.$$

Since in the subpopulation defined by any fixed $n \leq N$ the expression $Z_N - Z_n$ is independent of Z_n , we have

$$(11) \quad E_N e^{zn^t + (z_N - z_n)t} = E_N e^{zn^t} [\varphi(t)]^{N-n}.$$

From (10) and (11) we obtain the identity

$$(12) \quad P_N E_N \{e^{zn^t} [\varphi(t)]^{N-n}\} + (1 - P_N) E_N^* e^{z_N t} = [\varphi(t)]^N.$$

Dividing both sides by $[\varphi(t)]^N$ we obtain

$$(13) \quad P_N E_N \{e^{zn^t} [\varphi(t)]^{-n}\} + (1 - P_N) \frac{E_N^* e^{z_N t}}{[\varphi(t)]^N} = 1.$$

Let D'' be the subset of the complex plane in which $|\varphi(t)| \geq 1$ and denote by D the common part of the subsets D' and D'' . Since $\lim_{N \rightarrow \infty} (1 - P_N) = 0$, and since $|E_N^*(e^{z_N t})|$ is a bounded function of N , we have in D

$$(14) \quad \lim_{N \rightarrow \infty} (1 - P_N) \frac{E_N^* e^{z_N t}}{[\varphi(t)]^N} = 0.$$

Since

$$\lim_{N \rightarrow \infty} P_N E_N \{e^{zn^t} [\varphi(t)]^{-n}\} = E \{e^{zn^t} [\varphi(t)]^{-n}\},$$

we obtain from (13) and (14) the fundamental identity

$$(15) \quad E \{e^{zn^t} [\varphi(t)]^{-n}\} = 1,$$

for any point t in the set D .

4. Derivation of the probability that $Z_n \geq a$. In what follows in this and the subsequent sections we shall always assume that the random variable z satisfies the conditions I-IV of Lemma 2, even if this is not stated explicitly. Since it follows from Condition III that the set D' is the whole complex plane, we see that the identity (15) must hold for all points t for which $|\varphi(t)| \geq 1$.

Let $h_0 \neq 0$ be the real value for which $\varphi(h_0) = 1$. Substituting h_0 for t in (15) we obtain

$$(16) \quad E' e^{z_n h_0} = 1.$$

Let E_1 be the conditional expected value of $e^{z_n h_0}$ under the restriction that $Z_n \geq a$ and let E_0 be the conditional expected value of $e^{z_n h_0}$ under the restriction that $Z_n \leq b$. Furthermore denote $P(Z_n \geq a)$ by α . Then it follows from (16)

$$(17) \quad \alpha E_1 + (1 - \alpha) E_0 = 1.$$

Hence

$$(18) \quad \alpha = \frac{1 - E_0}{E_1 - E_0}.$$

If $h_0 > 0$ then $E_1 > 1$ and $E_0 < 1$. Hence (18) implies the inequality

$$(19) \quad \alpha \leq \frac{1}{E_1} \leq \frac{1}{e^{a h_0}}, \quad (h_0 > 0).$$

If $h_0 < 0$ then $E_1 < 1$ and $E_0 > 1$. Hence (18) implies the inequality

$$(20) \quad 1 - \alpha \leq \frac{1}{E_0} \leq \frac{1}{e^{b h_0}}, \quad (h_0 < 0).$$

We shall now derive lower and upper limits for E_0 and E_1 . We derive these limits under the assumption that $h_0 > 0$. To obtain a lower limit of E_0 consider a real variable ζ which is restricted to values > 1 . For any random variable u and any relation R we shall denote by $E(u | R)$ the conditional expected value of u under the restriction that R holds. Denote by $P(\zeta)$ the probability that $e^{h_0 Z_{n-1}} < \zeta e^{b h_0}$. Then we have

$$(21) \quad E_0 = \int_1^\infty \left\{ \zeta e^{b h_0} E \left[e^{h_0 z} \mid e^{h_0 z} \leq \frac{1}{\zeta} \right] \right\} dP(\zeta).$$

Hence a lower bound of E_0 is given by

$$(22) \quad E'_0 = e^{b h_0} \left\{ \text{g.l.b.}_\zeta \zeta E \left(e^{h_0 z} \mid e^{h_0 z} \leq \frac{1}{\zeta} \right) \right\},$$

where the symbol g.l.b._ζ stands for greatest lower bound with respect to ζ . Since $e^{b h_0}$ is an upper bound of E_0 , we obtain the limits

$$(23) \quad e^{b h_0} \left\{ \text{g.l.b.}_\zeta \zeta E \left(e^{h_0 z} \mid e^{h_0 z} \leq \frac{1}{\zeta} \right) \right\} \leq E_0 \leq e^{b h_0} \quad (h_0 > 0).$$

Let ρ be a real variable restricted to values > 0 and < 1 . Denote by $Q(\rho)$ the probability that $e^{h_0 Z_{n-1}} < \rho e^{ah_0}$. Then similarly to (21) we obtain

$$(24) \quad E_1 = \int_0^1 \left\{ \rho e^{ah_0} E \left(e^{h_0 z} \mid e^{h_0 z} \geq \frac{1}{\rho} \right) \right\} dQ(\rho).$$

Hence an upper bound of E_1 is given by

$$(25) \quad e^{ah_0} \left\{ \text{l.u.b.}_\rho \rho E \left(e^{h_0 z} \mid e^{h_0 z} \geq \frac{1}{\rho} \right) \right\}.$$

Since e^{ah_0} is a lower bound of E_1 , we obtain the following limits for E_1

$$(26) \quad e^{ah_0} \leq E_1 \leq e^{ah_0} \left\{ \text{l.u.b.}_\rho \rho E \left(e^{h_0 z} \mid e^{h_0 z} \geq \frac{1}{\rho} \right) \right\}, \quad (h_0 > 0).$$

In a similar way upper and lower limits can be derived for E_0 and E_1 when $h_0 < 0$. With the help of these limits upper and lower limits for α can be derived on the basis of equation (18). If $h_0 > 0$ then $E_1 > 1$, $E_0 < 1$ and consequently the right hand side of (18) is a monotonically decreasing function of E_0 and E_1 . Hence if E'_i is a lower, and E''_i is an upper bound of E_i ($i = 0, 1$), then

$$(27) \quad \frac{1 - E''_0}{E'_1 - E''_0} \leq \alpha \leq \frac{1 - E'_0}{E_1 - E'_0}, \quad (h_0 > 0),$$

In a similar way limits for α can be obtained when $h_0 < 0$. If both the absolute value of Ez and the variance of z are small, E_0 and E_1 will be nearly equal to e^{bh_0} and e^{ah_0} , respectively. Hence, in this case a good approximation to α is given by the expression

$$(28) \quad \bar{\alpha} = \frac{1 - e^{bh_0}}{e^{ah_0} - e^{bh_0}}.$$

The difference $\bar{\alpha} - \alpha$ approaches zero if both the mean and standard deviation of z converge to zero.

5. The characteristic function of n . Let Z_n be a random variable defined as follows: $Z_n = a$ if $Z_n \geq a$ and $Z_n = b$ if $Z_n \leq b$. Denote the difference $Z_n - Z_n$ by ϵ . Then ϵ is a random variable.

In what follows we shall neglect ϵ i.e., we shall substitute 0 for ϵ . No error is committed by doing so in the special case when z can take only two values d and $-d$ and the ratios a/d and b/d are integers, since in this case ϵ is exactly zero. Apart from this special case the variate ϵ will not be identical with the constant zero. However, the smaller the values $|Ez|$ and Ez^2 , the smaller the error we commit by neglecting ϵ . In fact, for arbitrary small positive numbers δ_1 and δ_2 the inequality $p(|\epsilon| \leq \delta_1) \geq 1 - \delta_2$ will hold if $|Ez|$ and Ez^2 are sufficiently small. Thus in the limiting case when Ez and Ez^2 approach zero the random variable ϵ reduces to the constant zero.

(a). *The characteristic function of n when only one of the quantities a and b is finite.* It will be sufficient to treat the case when a is finite and $b = -\infty$. In this case n is defined as the smallest positive integer for which $Z_n \geq a$. To make the probability of the existence of such a value n to be equal to 1 we have to assume that the expected value μ of z is positive. Since $b = -\infty$, the fundamental identity (15) need not hold for all points t of the set D . However, it follows easily from (13) that (15) holds for all points t in D whose real part is non-negative. Denote by $\psi(\tau)$ the characteristic function of n (τ is a purely imaginary variable). Since $Z_n = a$ (neglecting ϵ), and

$$E[\varphi(t)]^{-n} = \psi[-\log \varphi(t)],$$

identity (15) can be written as

$$(29) \quad e^{a\tau} \psi[-\log \varphi(t)] = 1.$$

Let $t(\tau)$ denote a root (with non-negative real part) of the equation in t

$$(30) \quad \log \varphi(t) + \tau = 0,$$

and substitute $t(\tau)$ for t in (29). Then we obtain

$$(31) \quad \psi(\tau) = e^{-a t(\tau)}.$$

As an illustration let us calculate $\psi(\tau)$ in the case when z is normally distributed. In this case

$$\log \varphi(t) = \mu t + \frac{\sigma^2}{2} t^2,$$

where μ is the mean and σ is the standard deviation of z . Hence

$$(32) \quad t(\tau) = \frac{-\mu \pm \sqrt{\mu^2 - 2\sigma^2\tau}}{\sigma^2}.$$

If we take the $+$ sign before the square root sign, the real part of $t(\tau)$ is non-negative, since the real part of $\sqrt{\mu^2 - 2\sigma^2\tau}$ is greater than or equal to μ . Hence the characteristic function of n is given by

$$(33) \quad \psi(\tau) = e^{-a/\sigma^2[-\mu + \sqrt{\mu^2 - 2\sigma^2\tau}]} \quad (\mu > 0).$$

(b). *The characteristic function of n when a and b both are finite.* Given the value of n , let p_n be the conditional probability that $Z_n = a$. Let p_n^* denote the probability that n is the smallest positive integer for which either $Z_n = a$ or $Z_n = b$ holds. Neglecting $Z_n - Z_n$, identity (15) can be written as

$$(34) \quad \sum_{n=1}^{\infty} [p_n e^{a\tau} + (1 - p_n) e^{b\tau}] [\varphi(t)]^{-n} p_n^* = 1.$$

Let $\psi_1(\tau)$ be the characteristic function of n in the subpopulation where $Z_n = a$, and let $\psi_2(\tau)$ be the characteristic function of n in the subpopulation where $Z_n =$

b. Furthermore let $\psi(\tau)$ be the characteristic function of n in the total population.

Since we neglect the difference $\bar{Z}_n - Z_n$, it follows from (18) that the probability α that $\bar{Z}_n = a$ is given by

$$(35) \quad \alpha = \frac{1 - e^{bh_0}}{e^{ah_0} - e^{bh_0}}.$$

Putting $1 - p_n = q_n$ the following relations hold

$$(36) \quad \psi_1[-\log \varphi(t)] = \frac{\sum_{n=1}^{\infty} p_n p_n^* [\varphi(t)]^{-n}}{\sum p_n p_n^*} = \frac{\sum p_n p_n^* [\varphi(t)]^{-n}}{\alpha}$$

$$(37) \quad \psi_2[-\log \varphi(t)] = \frac{\sum q_n p_n^* [\varphi(t)]^{-n}}{\sum q_n p_n^*} = \frac{\sum q_n p_n^* [\varphi(t)]^{-n}}{1 - \alpha}$$

$$(38) \quad \begin{aligned} \psi[-\log \varphi(t)] &= \sum p_n^* [\varphi(t)]^{-n} = \sum (p_n + q_n) [\varphi(t)]^{-n} p_n^* \\ &= \alpha \psi_1[-\log \varphi(t)] + (1 - \alpha) \psi_2[-\log \varphi(t)]. \end{aligned}$$

Putting $-\log \varphi(t) = \tau$ we obtain from (34), (36) and (37)

$$(39) \quad \alpha \psi_1(\tau) e^{a\tau} + (1 - \alpha) \psi_2(\tau) e^{b\tau} = 1.$$

According to Lemma 2 the equation $-\log \varphi(t) = 0$ has two different real roots in t , $t = 0$ and $t = h_0$, and $\varphi'(0)$ and $\varphi'(h_0)$ both are unequal to zero. Hence, if $\varphi(t)$ is not singular at $t = 0$ and $t = h_0$, the equation

$$-\log \varphi(t) = \tau,$$

has two roots $t_1(\tau)$ and $t_2(\tau)$ for sufficiently small values of τ such that $\lim_{\tau \rightarrow 0} t_1(\tau) = 0$ and $\lim_{\tau \rightarrow 0} t_2(\tau) = h_0$. Since the identity (15) holds for all values of t for which $|\varphi(t)| \geq 1$, and since $|\varphi[t_1(\tau)]| = |\varphi[t_2(\tau)]| = 1$ for all imaginary values of τ , it follows from (39) that both equations hold

$$(39') \quad \alpha \psi_1(\tau) e^{a t_1(\tau)} + (1 - \alpha) \psi_2(\tau) e^{b t_1(\tau)} = 1,$$

$$(39'') \quad \alpha \psi_1(\tau) e^{a t_2(\tau)} + (1 - \alpha) \psi_2(\tau) e^{b t_2(\tau)} = 1.$$

Solving these two linear equations we obtain $\psi_1(\tau)$ and $\psi_2(\tau)$. The characteristic function $\psi(\tau)$ is given by

$$\psi(\tau) = \alpha \psi_1(\tau) + (1 - \alpha) \psi_2(\tau).$$

As an illustration we shall determine $\psi_1(\tau)$, $\psi_2(\tau)$ and $\psi(\tau)$ when z has a normal distribution with mean μ and standard deviation σ . We have

$$-\log \varphi(t) = -\mu t - \frac{\sigma^2}{2} t^2 = \tau.$$

Hence

$$(40) \quad t = \frac{-\mu \pm \sqrt{\mu^2 - 2\sigma^2\tau}}{\sigma^2}.$$

Putting $e^a = A$ and $e^b = B$ we obtain from (39) and (40)

$$(41) \quad \alpha\psi_1(\tau)A^{-\mu/\sigma^2+1/\sigma^2\sqrt{\mu^2-2\sigma^2\tau}} + (1-\alpha)\psi_2(\tau)B^{-\mu/\sigma^2+1/\sigma^2\sqrt{\mu^2-2\sigma^2\tau}} = 1$$

$$(42) \quad \alpha\psi_1(\tau)A^{-\mu/\sigma^2-1/\sigma^2\sqrt{\mu^2-2\sigma^2\tau}} + (1-\alpha)\psi_2(\tau)B^{-\mu/\sigma^2-1/\sigma^2\sqrt{\mu^2-2\sigma^2\tau}} = 1.$$

These two equations are valid for any imaginary value of τ . Since $h_0 = \frac{-2\mu}{\sigma^2}$, we obtain from (35)

$$(43) \quad \alpha = \frac{1 - B^{-2\mu/\sigma^2}}{A^{-2\mu/\sigma^2} - B^{-2\mu/\sigma^2}}.$$

Let

$$(44) \quad g_1 = -\frac{\mu}{\sigma^2} + \frac{1}{\sigma^2} \sqrt{\mu^2 - 2\sigma^2\tau},$$

and

$$(45) \quad g_2 = -\frac{\mu}{\sigma^2} - \frac{1}{\sigma^2} \sqrt{\mu^2 - 2\sigma^2\tau}.$$

Then we obtain from (41) and (42)

$$(46) \quad \alpha\psi_1(\tau) = \frac{B^{g_1} - B^{g_2}}{A^{g_1}B^{g_2} - A^{g_2}B^{g_1}},$$

and

$$(47) \quad (1-\alpha)\psi_2(\tau) = \frac{A^{g_1} - A^{g_2}}{A^{g_1}B^{g_2} - A^{g_2}B^{g_1}}.$$

Hence the characteristic function of n is given by

$$(48) \quad \psi(\tau) = \frac{A^{g_1} + B^{g_1} - A^{g_2} - B^{g_2}}{A^{g_1}B^{g_2} - A^{g_2}B^{g_1}}.$$

6. The distribution of n when z is normally distributed. (a) *The case when a is finite and $b = -\infty$.* In this case the characteristic function of n is given by (33). Let

$$(49) \quad m = \frac{\mu^2}{2\sigma^2} n.$$

Then the characteristic function of m is given by

$$(50) \quad \psi^*(t) = e^{e^{(1-\sqrt{1-i})}},$$

where

$$(51) \quad c = \frac{a\mu}{\sigma^2} > 0.$$

The distribution of m is given by

$$(52) \quad \frac{1}{2\pi i} \int_{-\infty}^{\infty} e^{c(1-\sqrt{1-t})-mt} dt.$$

Let

$$(53) \quad G(c, m) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} e^{-c\sqrt{1-t}-mt} dt,$$

and

$$(54) \quad H(c, m) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{1}{\sqrt{1-t}} e^{-c\sqrt{1-t}-mt} dt.$$

Since

$$(55) \quad \frac{1}{2\pi i} \frac{d}{dt} e^{-c\sqrt{1-t}-mt} = \frac{1}{2\pi i} \left(\frac{c}{2\sqrt{1-t}} - m \right) e^{-c\sqrt{1-t}-mt}$$

we have

$$(56) \quad \frac{c}{2} H(c, m) - mG(c, m) = \frac{1}{2\pi i} [e^{-c\sqrt{1-t}-mt}]_{-\infty}^{\infty} = 0.$$

From (53) and (54) we obtain

$$(57) \quad \frac{\partial H(c, m)}{\partial c} + G(c, m) = 0.$$

From (56) and (57) it follows that

$$(58) \quad \frac{c}{2} H(c, m) + m \frac{\partial H(c, m)}{\partial c} = 0.$$

Hence

$$(59) \quad \log H(c, m) = -\frac{c^2}{4m} + \log \lambda(m)$$

where $\lambda(m)$ is some function of m only. Thus

$$(60) \quad H(c, m) = \lambda(m) e^{-c^2/4m}$$

Now we shall determine $\lambda(m)$. We have

$$(61) \quad \lambda(m) = H(0, m) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{1}{\sqrt{1-t}} e^{-mt} dt.$$

Since $(1 - t)^{-1/2}$ is the characteristic function of $\frac{1}{2}\chi^2$ where χ^2 has the χ^2 -distribution with one degree of freedom, the right hand side of (61) is equal to

$$\frac{1}{\Gamma(\frac{1}{2})\sqrt{m}} e^{-m}.$$

Hence

$$(62) \quad \lambda(m) = \frac{1}{\Gamma(\frac{1}{2})\sqrt{m}} e^{-m}.$$

From (60) and (61) we obtain

$$(63) \quad H(c, m) = \frac{1}{\Gamma(\frac{1}{2})\sqrt{m}} e^{-c^2/4m-m}.$$

From (56) and (63) we obtain

$$(64) \quad G(c, m) = \frac{c}{2\Gamma(\frac{1}{2})m^{3/2}} e^{-c^2/4m-m}.$$

Hence the distribution of m is given by

$$(65) \quad F(m) dm = \frac{c}{2\Gamma(\frac{1}{2})m^{3/2}} e^{-c^2/4m-m+c} dm, \quad (0 \leq m < \infty).$$

Let $m = \frac{c}{2} m^*$. Then the distribution of m^* is given by

$$(66) \quad \begin{aligned} D(m^*) dm^* &= \frac{c^2/2}{2\Gamma(\frac{1}{2})\left(\frac{c}{2}\right)^{3/2} (m^*)^{3/2}} e^{-(c/2)(1/m^* + m^* - 2)} dm^* \\ &= \frac{\sqrt{c}}{\sqrt{2\pi}(m^*)^{3/2}} e^{-(c/2)(1/m^* + m^* - 2)} dm^*. \end{aligned}$$

The function $\frac{1}{m^*} + m^* - 2$ is non-negative and is equal to zero only when $m^* = 1$.

If c is large, then $D(m^*)$ is exceedingly small for values of m^* not close to 1.

Expanding $\frac{1}{m^*} + m^* - 2$ in a Taylor series around $m^* = 1$, we obtain

$$(67) \quad \frac{1}{m^*} + m^* - 2 = (m^* - 1)^2 + \text{higher order terms}.$$

Hence for large c

$$(68) \quad D(m^*) dm^* \sim \frac{\sqrt{c}}{\sqrt{2\pi}} e^{-(c/2)(m^*-1)^2} dm^*,$$

i.e., if c is large m^* is nearly normally distributed with mean equal to 1 and standard deviation $\frac{1}{\sqrt{c}}$.

(b). *The case when a and b both are finite.* In this case the characteristic function of n is given by (48). Let

$$m = \frac{\mu^2}{2\sigma^2} n \quad \text{and} \quad d = -\frac{\mu}{\sigma^2}.$$

Then the characteristic function of m is given by

$$(69) \quad \psi^*(t) = \frac{A^{h_1} + B^{h_2} - A^{h_2} - B^{h_1}}{A^{h_1} B^{h_2} - A^{h_2} B^{h_1}},$$

where

$$(70) \quad h_1 = d(1 - \sqrt{1-t}), \quad h_2 = d(1 + \sqrt{1-t}),$$

and t is an imaginary variable. Putting $A^d = \bar{A}$, $B^d = \bar{B}$, $da = \bar{a}$ and $db = \bar{b}$, the characteristic function of m can be written as

$$(71) \quad \begin{aligned} \psi^*(t) &= \frac{\bar{A}(e^{-\bar{a}\sqrt{1-t}} - e^{\bar{a}\sqrt{1-t}}) + \bar{B}(e^{\bar{b}\sqrt{1-t}} - e^{-\bar{b}\sqrt{1-t}})}{\bar{A}\bar{B}(e^{(\bar{b}-\bar{a})\sqrt{1-t}} - e^{(\bar{a}-\bar{b})\sqrt{1-t}})} \\ &= \frac{\bar{A}(e^{-\bar{b}\sqrt{1-t}} - e^{(2\bar{a}-\bar{b})\sqrt{1-t}}) + \bar{B}(e^{\bar{a}\sqrt{1-t}} - e^{(\bar{a}-2\bar{b})\sqrt{1-t}})}{\bar{A}\bar{B}(1 - e^{2(\bar{a}-\bar{b})\sqrt{1-t}})}. \end{aligned}$$

It will be sufficient to consider only the case when $\mu > 0$, since the case < 0 can be treated in a similar way. Then $\bar{a} < 0$ and $\bar{b} > 0$. Since the real part of $+\sqrt{1-t}$ is greater than or equal to one, we have

$$(72) \quad |e^{2(\bar{a}-\bar{b})\sqrt{1-t}}| < 1,$$

for any imaginary value of t . Let

$$(73) \quad T = e^{2(\bar{a}-\bar{b})\sqrt{1-t}}.$$

Then

$$(74) \quad \frac{1}{1-T} = \sum_{j=0}^{\infty} T^j.$$

From (71) and (74) it follows that $\psi^*(t)$ can be written in the form of an infinite series.

$$(75) \quad \psi^*(t) = \sum_{i=1}^{\infty} r_i e^{-\lambda_i \sqrt{1-t}},$$

where λ_i and r_i are constants and $\lambda_i > 0$. Each term of this series is a characteristic function of the form given in (50) except for a proportionality factor. Let $F_i(m)$ be the distribution of m corresponding to the characteristic function $e^{-\lambda_i \sqrt{1-t}}$. Then $F_i(m)$ can be obtained from (65) by substituting λ_i for c . Since we may integrate the right hand side member of (75) term by term, the distribution of m is given by

$$(76) \quad F(m) dm = \left(\sum_{i=1}^{\infty} \frac{r_i}{e^{\lambda_i}} F_i(m) \right) dm.$$

Since m is a discrete variable, it may seem paradoxical that we obtained a probability density function for m . However, the explanation lies in the fact that we neglected $\epsilon = Z_n - Z_n^*$ and this quantity is zero only in the limiting case when μ and σ approach zero.

If $|\mu|$ and σ are sufficiently small as compared with a and $|b|$, the distribution of m given in (76) will be a good approximation to the exact distribution of m , even if z is not normally distributed. The reason for this can be indicated as follows: Let

$$(77) \quad z_i^* = \sum_{j=-(i-1)r+1}^{ir} z_j, \quad (i = 1, 2, \dots, \text{ad inf.})$$

where r is a given positive integer. Since the variates z_j are independently distributed each having the same distribution, under some weak conditions the variates z_i^* ($i = 1, 2, \dots, \text{ad inf.}$) will be nearly normally distributed for large r . Hence, considering the cumulative sums $Z_i^* = z_1^* + z_2^* + \dots + z_i^*$ ($i = 1, 2, \dots, \text{ad inf.}$), the distribution given in (76) is applicable with good approximation, provided that $r|\mu|$ and $\sqrt{r}\sigma$ are small as compared with a and $|b|$ so that the difference $\epsilon^* = Z_n^* - Z_n^*$ can be neglected.

7. The exact probability distribution of Z_n and the exact characteristic function of n when z can take only integral multiples of a given constant d . In the previous sections we derived the probability $P(Z_n \geq a)$ and the characteristic function of n under the assumption that the quantity by which Z_n may differ from a or b is small and can be neglected. This can be done whenever $|Ez|$ and Ez^2 are small. However, if $|Ez|$ or Ez^2 is not small, it is desirable to derive the exact probability distribution of Z_n and the exact characteristic function of n . Both are obtained in the present section for random variables z which can take only a finite number of integral multiples of a given constant d . This is a rather general result, since any distribution of z can be approximated arbitrarily fine by a discrete distribution of the above type if the constant d is chosen sufficiently small.

There is no loss of generality in assuming that $d = 1$, since the quantity d can be chosen as the unit of measurement. Thus, we shall assume that z takes only a finite number of integral values. Let g_1 and g_2 be two positive integers such that $P(z = -g_1)$ and $P(z = g_2)$ are positive and z can take only integral values $\geq -g_1$ and $\leq g_2$. Denote $P(z = i)$ by h_i . Then the characteristic function of z is given by

$$(78) \quad \varphi(t) = \sum_{i=-g_1}^{g_2} h_i e^{it}.$$

To obtain the roots of the equation $\varphi(t) = 1$, we put $e^t = u$ and solve the equation

$$(79) \quad \sum_{i=-g_1}^{g_2} h_i u^i = 1.$$

Denote $g_1 + g_2$ by g and let the g roots of (79) be u_1, \dots, u_g , respectively. We shall assume that no two roots are equal, i.e., $u_i \neq u_j$ for $i \neq j$. Substituting u_i for e^t in the identity (15) we obtain

$$(80) \quad E(u_i^{Z_n}) = 1 \quad (i = 1, \dots, g).$$

Denote by $[a]$ the smallest integer $\geq a$, and by $[b]$ the largest integer $\leq b$. Then Z_n can take only the values

$$(81) \quad [b] - g_1 + 1, \quad [b] - g_1 + 2, \dots, [b], [a], \quad [a] + 1, \dots, [a] + g_2 - 1.$$

Denote the g different integers in (81) by c_1, \dots, c_g , respectively. Furthermore, denote $P(Z_n = c_i)$ by ξ_i . Then equations (80) can be written as

$$(82) \quad \sum_{i=1}^g \xi_i u_i^{c_i} = 1 \quad (i = 1, \dots, g).$$

Let Δ be the determinant value of the matrix $\|u_i^{c_j}\|$ ($i, j = 1, \dots, g$) and let Δ_j be the determinant we obtain from Δ by substituting 1 for the elements in the j th column. If $\Delta \neq 0$, it follows from (82) that $P(Z_n = c_j) = \xi_j$ is given by

$$(83) \quad \xi_j = \frac{\Delta_j}{\Delta}.$$

Hence, $P(Z_n \geq a) = \sum_j (\Delta_j / \Delta)$ summed for all values of j for which $c_j \geq a$.

From the probability distribution of Z_n we can easily derive the expected value En of n . In fact, differentiating the fundamental identity (15) with respect to t at $t = 0$ we obtain

$$(84) \quad E \left[Z_n - \frac{\varphi'(0)}{\varphi(0)} n \right] = 0.$$

Since $\frac{\varphi'(0)}{\varphi(0)} = Ez$, we obtain from (84)

$$(85) \quad En = \frac{EZ_n}{Ez} = \frac{1}{Ez} \sum_{i=1}^g c_i \Delta_i.$$

Now we shall derive the exact characteristic function of n . Denote by $\psi_i(\tau)$ (τ is a purely imaginary variable) the characteristic function of the conditional distribution of n under the restriction that $Z_n = c_i$. Let $t_1(\tau), \dots, t_g(\tau)$ be g roots of the equation

$$(86) \quad \varphi(t) = e^{-\tau},$$

such that

$$(87) \quad \lim_{\tau \rightarrow 0} e^{t_i(\tau)} = u_i.$$

Substituting $t_i(\tau)$ for t in the fundamental identity (15) we obtain

$$(88) \quad \sum_{j=1}^g \xi_j e^{t_j(\tau)} \psi_j(\tau) = 1 \quad (i = 1, \dots, g).$$

These equations are linear in the unknowns $\psi_1(\tau), \dots, \psi_g(\tau)$ and the determinant of these equations is given by

$$(89) \quad \delta(\tau) = \begin{vmatrix} \xi_1 e^{c_{11} t_1(\tau)} & \dots & \xi_g e^{c_{g1} t_1(\tau)} \\ \xi_1 e^{c_{12} t_2(\tau)} & \dots & \xi_g e^{c_{g2} t_2(\tau)} \\ \vdots & \dots & \vdots \\ \xi_1 e^{c_{1g} t_g(\tau)} & \dots & \xi_g e^{c_{gg} t_g(\tau)} \end{vmatrix}.$$

Obviously, $\delta(0) = \xi_1 \xi_2 \dots \xi_g \Delta$. Hence if $\xi_i \neq 0$ ($i = 1, \dots, g$) and $\Delta \neq 0$, also $\delta(0) \neq 0$ and consequently $\delta(\tau) \neq 0$ for any τ with sufficiently small absolute value. Thus, $\psi_1(\tau), \dots, \psi_g(\tau)$ can be obtained by solving the linear equations (88). The characteristic function $\psi(\tau)$ of the unconditional distribution of n is given by

$$(90) \quad \psi(\tau) = \sum_{i=1}^g \xi_i \psi_i(\tau).$$

SOME IMPROVEMENTS IN WEIGHING AND OTHER EXPERIMENTAL TECHNIQUES¹

BY HAROLD HOTELLING

Columbia University

When several quantities are to be ascertained there is frequently an opportunity to increase the accuracy and reduce the cost by combining suitably in one experiment what might ordinarily be considered separate operations. The theory of design of experiments developed as a branch of modern mathematical statistics, and of which fundamental considerations are set forth in R. A. Fisher's book [1], provides many improvements of this kind. Since the main interests of Fisher and other originators of this theory have been in biology, the applications so far made have been chiefly biological in character, excepting for certain economic and social investigations involving stratified sampling. The possibilities of improvement of physical and chemical investigations through designed experiments based on the theory of statistical inference have scarcely begun to be explored.

The following example is due to F. Yates [2]. A chemist has seven light objects to weigh, and the scale also requires a zero correction, so that eight weighings are necessary. The standard error of each weighing is denoted by σ , the variance therefore by σ^2 . Since the weight assigned to each object by customary techniques is the difference between the reading of the scale when carrying that object and when empty, the variance of the assigned weight is $2\sigma^2$, and its standard error is $\sigma\sqrt{2}$.

The improved technique suggested by Yates consists of weighing all seven objects together, and also weighing them in groups of three so chosen that each object is weighed four times altogether, twice with any other object and twice without it. Calling the readings from the scale y_1, \dots, y_8 we then have as equations for determining the unknown weights a, b, \dots, g ,

$$\begin{array}{rcl} a + b + c + d + e + f + g & = & y_1 \\ a + b + c & & = y_2 \\ a & + d + e & = y_3 \\ a & & + f + g = y_4 \\ b & + d & + f = y_5 \\ b & & + e + g = y_6 \\ c + d & & + g = y_7 \\ c & + e + f & = y_8. \end{array}$$

¹ Presented at the Wellesley meeting of the Institute of Mathematical Statistics, Aug. 13, 1944.

Any particular weight is found by adding together the four equations containing it, subtracting the other four, and dividing by 4. Thus

$$a = \frac{y_1 + y_2 + y_3 + y_4 - y_5 - y_6 - y_7 - y_8}{4}.$$

The variance of a sum of independent observations is the sum of their variances, as is well known, and the variance of c times an observation is c^2 times the variance of that observation. Taking $c = \frac{1}{4}$ for the first four terms in the expression for a and $c = -\frac{1}{4}$ for the others gives for the variance of a by this method $\sigma^2/2$, which is only one-fourth that for the direct method. The standard error, or probable error, has been halved. If a degree of accuracy is required calling for repetition a certain number of times of the weighings by the direct method, then only one-fourth as many weighings are needed by Yates' method to procure the same accuracy in the average.

A further improvement, which does not seem to have been mentioned in the literature, will be obtained if Yates' procedure is modified by placing in the other pan of the scale those of the objects not included in one of his weighings. Calling the readings in this case z_1, \dots, z_8 , we have

$$\begin{aligned} a + b + c + d + e + f + g &= z_1 \\ a + b + c - d - e - f - g &= z_2 \\ a - b - c + d + e - f - g &= z_3 \\ a - b - c - d - e + f + g &= z_4 \\ -a + b - c + d - e + f - g &= z_5 \\ -a + b - c - d + e - f + g &= z_6 \\ -a - b + c + d - e - f + g &= z_7 \\ -a - b + c - d + e + f - g &= z_8. \end{aligned}$$

From these equations,

$$a = \frac{z_1 + z_2 + z_3 + z_4 - z_5 - z_6 - z_7 - z_8}{8},$$

with a like expression for each of the other unknowns. The variance of each unknown by this method is $\sigma^2/8$. The standard error is half that by Yates' method, or a quarter of its value by the direct method of weighing each object separately. The number of repetitions required to procure a particular standard error in the mean is one-sixteenth that by the direct method.

A simpler example illustrating the same point is that of two objects to be weighed, with a scale already corrected for bias. Again let σ^2 be the variance of an individual weighing. If we weigh the two objects together in one pan of the scale, and then in opposite pans, we have as equations for the unknown weights a and b ,

$$a + b = z_1, \quad a - b = z_2,$$

whence

$$a = (z_1 + z_2)/2, \quad b = (z_1 - z_2)/2.$$

The variances of a and b by this method are both equal to $\sigma^2/2$, half the value when the two objects are weighed separately. The means found from a number of pairs of weighings of sums and differences have the same precision as those found from twice as many pairs of weighings of the objects separately.

Further economies of effort, or gains in accuracy, are possible with larger numbers of weighings and of objects to be weighed. These improvements can to some extent be applied also to other types of measurement, as of distances, since it is sometimes possible to measure the sum of a number of such quantities, or the difference between two such sums, with approximately the same accuracy as a single one of them. The outstanding case, however, seems to be that of weighing on a balance objects light enough so that their aggregate weight is below the maximum for which the balance was designed, since in this case it is quite reasonable to assume that the several recorded results all have the same standard error σ and that they are independent.

In what follows, some principles underlying the design of efficient schemes of this kind will be developed and applied to obtain some additional plans. However no comprehensive general solution has been reached; this appears to be a matter for further mathematical research. Also, we leave aside in this paper the problem of estimating the error variance. All this discussion is based on the minimization of the actual variance. In order to utilize the results it is necessary that this variance be either known a priori or estimated from the residuals from the least-square solution. The latter type of estimate is in some ways more satisfactory, since it refers to the actual experiment rather than to some previous experiments which may not have been made under exactly the same conditions. But in order to have such an estimate it is necessary that the number of observations exceed the number of unknowns, and desirable that the excess shall have a large enough value to insure a stable estimate of the error variance σ^2 . The appropriate test for significance, or determination of confidence limits for the unknowns, must then utilize the Student distribution or its generalization, the variance ratio distribution, which take full account of the instability caused by an inadequate number of degrees of freedom for estimating σ .

It is only when σ is known exactly apart from the experiment being designed that the criteria we here consider are exactly applicable. In other cases there may need to be a balancing, in the design of the experiment, between the desiderata of *minimum* variance and of *accurately known* variance, with the accuracy of this knowledge depending on the number of available degrees of freedom. A theory of design taking full account of this consideration would require a use of the power functions of the Student distribution and the variance ratio distribution, discovered respectively by R. A. Fisher [3] and P. C. Tang [4].

We shall denote by N the number of weighings to be made, and by p the number of objects to be weighed. In order that it be possible to determine the un-

known weights from the observations it is necessary that $p \leq N$, and if a possible bias in the scale must be eliminated by means of the same data it is necessary that $p \leq N - 1$. Supposing these conditions to be satisfied, we shall show, among other things, that the minimum possible variance for one of the unknowns is σ^2/N ; that the experiment may be arranged so that a selected one of the unknowns has exactly this minimum variance excepting when N is odd and a bias must be allowed for also; and that for some, but not all, combinations of p and N , this minimum variance is attained for all the unknowns simultaneously. This minimum value σ^2/N is of course equal to the variance of the mean of N weighings of one object alone, disregarding the rest; but it will be seen below that by complex experiments of the kind indicated, determinations from the same number of weighings of the other weights also can at the same time be made with some finite variance, which may or may not have the minimum value.

The following notation will be used in the proof. Let $x_{i\alpha} = 1$ or -1 if the i th object is included in the α th weighing by being placed respectively in the left- or right-hand pan, and let $x_{i\alpha} = 0$ if the i th object is not included in the α th weighing. Here $i = 1, 2, \dots, p$ and $\alpha = 1, \dots, N$. Let y_α be the result recorded for the α th weighing, let Δ_α be the error in this result, and let b_i be the true weight of the i th object, so that we have the N equations

$$(1) \quad x_{1\alpha}b_1 + x_{2\alpha}b_2 + \dots + x_{p\alpha}b_p = y_\alpha + \Delta_\alpha,$$

provided there is no bias, or if by y_α we mean the observed weight corrected for a bias known a priori. Under these conditions the estimate of each of the b_i 's having the properties of zero bias and minimum variance is that provided by the method of least squares. This statement, which does not depend on any assumption of a normal or other particular form of distribution of the errors, has been known long but not widely, since there is an easier derivation of the method by the application to the normal distribution to the method of maximum likelihood. Its proof, due originally to Laplace, has appeared in many forms in the work of Gauss and later authors [5]; the latest version is by the present writer [6].

Letting S stand for summation over all the N weighings we put

$$(2) \quad a_{ij} = Sx_{i\alpha}x_{j\alpha}, \quad g_i = Sx_{i\alpha}y_\alpha,$$

and write the normal equations in the form

$$\Sigma a_{ij}b_j = g_i,$$

where Σ stands for a sum with respect to j from 1 to p . From the usual theory of least squares (cf. for example the reference last cited) it is known that the standard error of the determination of b_i from these equations—which is the

minimum possible standard error of b_1 for any way of combining the observations—is σ times the square root of A_{11}/A , where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix},$$

and A_{11} is the minor of A obtained by deleting the first row and column.

The matrices of A and of A_{11} are known to be positive definite or semi-definite. The semi-definite case is excluded by the consideration that the normal equations shall actually determine the unknowns. Hence the inverse of the latter matrix exists and is positive definite. But this inverse, which we may write

$$d = \begin{bmatrix} d_{22} & \cdots & d_{2p} \\ \cdot & \cdot & \cdot \\ d_{p2} & \cdots & d_{pp} \end{bmatrix},$$

consists of the coefficients in the identity

$$A/A_{11} = a_{11} - \sum_{i,j=2}^p d_{ij} a_{i1} a_{j1}.$$

which is obtained by expanding A with reference to its first row and first column. The positive definite character of d therefore leads to the following

LEMMA: If a_{12}, \dots, a_{1p} ($= a_{21}, \dots, a_{p1}$ respectively) are free to vary while the other elements of A remain fixed, the maximum value of A/A_{11} is a_{11} , and is attained when and only when $a_{12} = a_{13} = \dots = a_{1p} = 0$.

From this it is evident that the variance of b_1 , namely $\sigma^2 A_{11}/A$, cannot be less than σ^2/a_{11} , and will reach this value only if the experiment is so arranged that the elements after the first in the first row and column of A are all zero. That such an arrangement is possible may be seen by a consideration of the matrix

$$X = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{p1} \\ x_{12} & \cdot & \cdot & \cdot & x_{p2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{1N} & \cdot & \cdot & \cdot & x_{pN} \end{bmatrix}$$

whose elements are restricted to be 1's, -1's and 0's. The condition $a_{11} = 0$, by (2), means simply that $Sx_{i\alpha}x_{1\alpha} = 0$, a condition which may be expressed by saying that the first column of X is orthogonal to the i th column. The condition that the variance of b_1 have its minimum value σ^2/a_{11} is thus, according to the lemma, that the first column of X shall be orthogonal to all the others. The *minimum minimorum* of this variance will be reached if the first row of X is

not only orthogonal to all the others, but consists entirely of 1's and -1's, so that $a_{11} = N$. The value of this minimum minimorum is σ^2/N .

If there is a possible bias b_0 this procedure needs to be modified by the addition of b_0 to the left member of (1) and subsequent treatment of this term like the others, putting $x_{0\alpha} = 1$ in (2), and modifying X by adjoining a column of 1's. The necessary and sufficient condition that the variance of b_1 shall equal σ^2/N is then that the column

$$x_{11}$$

$$x_{12}$$

$$\dots$$

$$x_{1N}$$

shall consist entirely of 1's and -1's and shall be orthogonal to a column consisting of 1's, and to all the other columns of X .

If no bias needs to be eliminated the experiment can be arranged so that the variance of b_1 is σ^2/N merely by filling up the first column of X with 1's and -1's in any arbitrary manner, and then choosing the later columns so as to be orthogonal to this first one, and so that all are linearly independent. This can be accomplished, for example by choosing the first element in all the columns to be the same as that in the first column; choosing the i th element in the i th column ($i = 2, 3, \dots, p$) to be the negative of the i th element in the first column; and making all the other elements of X zero.

When a bias is to be eliminated, so that there is a column of 1's in X corresponding to b_0 , it is necessary that N be even in order that the column of X corresponding to b_1 may consist of 1's and -1's in equal numbers, without any 0's, a condition essential for the orthogonality between these two columns with the maximum value N for a_{11} . Supposing N even, let us assign the value 1 to each of the first $N/2$ elements of the column corresponding to b_1 and the value -1 to the last $N/2$ elements of this row. The remaining rows of X may then be filled up by the same method as that indicated above for the case in which there is no bias. The variance of b_1 will then take its theoretical minimum value σ^2/N .

If N is odd and there is a possible bias, the column of X corresponding to b_1 can be filled up with 1's and -1's in equal numbers, with a single zero, and the remaining columns can be made orthogonal to it. The variance of b_1 in this case will be $\sigma^2/(N-1)$.

The method suggested above for filling up the later columns of X is convenient for the proof, but is not usually to be recommended in practice, since other methods will in all but the simplest cases give smaller standard errors for the unknowns other than the first. For some values of N and p it is possible to determine all the unknowns with equal and minimum variance. These are the cases in which all the columns of X can be made mutually orthogonal and

without zeros, excepting that the column corresponding to b_0 may contain some zeros. Thus for $N = 4$ the scheme of weighing represented by the matrix

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

whose columns are all mutually orthogonal, may be applied to weigh three objects when there is a possible bias, or four where there is not, with variance $\sigma^2/4$ for each of the unknowns in either case. The matrix $X'X$ of the normal equations has the form

$$\begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$

Calling the results of the weighings y_0, y_1, y_2, y_3 in the case of possible bias we have for the unknowns the expressions

$$b_1 = (y_0 + y_1 - y_2 - y_3)/4$$

$$b_2 = (y_0 - y_1 + y_2 - y_3)/4$$

$$b_3 = (y_0 - y_1 - y_2 + y_3)/4.$$

The complete orthogonality exemplified by this design has several advantages besides the fact that the variance of each of the unknowns has the same minimum value as if all the weighings were to be devoted to it alone (or half the value of the variance of this unknown if half the weighings were devoted to it plus bias and half to determining the bias). The diagonal form of the matrix $X'X$ means that the labor of solving normal equations, which is sometimes formidable, is reduced to the trivial task of dividing by N . Also, the diagonal form of this matrix implies that its inverse is also of diagonal form, from which it follows that the estimates of the different unknowns are statistically independent. Consequently the variances, or standard errors, of linear functions of the unknowns are easy to find. Thus the variance of the difference between the estimates of two of the weights is simply the sum of their variances. But of course if the main object of the experiment is to determine a particular difference of this kind, or any other linear function of the weights, a different design should be sought to minimize the particular variance which is of interest.

In contrast to the satisfactory design possible with four weighings, no complete orthogonality is possible with six weighings, or with any odd number, if the number of objects to be weighed is the maximum possible for the number of

weighings and if each object is actually to enter into each weighing in one pan or the other. For $N = 3$ and bias known to be zero consider the scheme

$$X = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix},$$

which corresponds to weighing two objects, first together in one pan, then in opposite pans, and then weighing one alone. Calling b_1 the weight of the object that has been on the scale through all three weighings and b_2 the other we have the estimates

$$b_1 = (y_1 + y_2 + y_3)/3$$

$$b_2 = (y_1 - y_2)/2,$$

with respective variances

$$\sigma_1^2 = \sigma^2/3, \quad \sigma_2^2 = \sigma^2/2.$$

Thus the first weight is determined with the minimum possible variance but the second is not.

An alternative method of weighing under these same conditions is to weigh both objects in one pan together twice and to weigh them in opposite pans once. This gives

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix},$$

with the normal equations

$$3b_1 + b_2 = y_1 + y_2 + y_3$$

$$b_1 + 3b_2 = y_1 + y_2 - y_3,$$

whose solution is

$$b_1 = (y_1 + y_2 + 2y_3)/4$$

$$b_2 = (y_1 + y_2 - 2y_3)/4,$$

and variances

$$\sigma_1^2 = \sigma_2^2 = \frac{3}{4}\sigma^2.$$

Thus the weights are by this method determined with equal accuracy, which is better than by the preceding method for one of the objects but worse for the other. To choose between the two methods it is therefore appropriate to take into consideration the relative accuracy desired in the weights of the two objects. Either method is better than weighing the objects separately.

Either of these two X matrices can also be made the basis for weighing a single

object when the scale is suspected of having a bias. The weight of this object will be estimated as b_2 , and will have the variance $\frac{1}{2}\sigma^2$ by the first method, or $\frac{2}{3}\sigma^2$ by the second. Thus the second method is distinctly superior in this case.

Orthogonality between columns obviously requires both negative and positive signs, corresponding to weighings in both pans of the balance. Thus the experimental designs of maximum efficiency for weighing on a balance are not available with a spring scale, or in making measurements of any kind in which it is not possible to arrange that the quantities read off are differences. In such cases the elements of X are restricted to be 1 or 0. Let us now consider some of the simplest cases of this kind, assuming for simplicity that $\sigma = 1$. We shall deal only with cases in which there is no bias.

For $N = 3$, $p = 2$ the simple experiment of weighing one object twice and the other once yields variances $\frac{1}{2}$ and 1 respectively. All other designs are in this case less satisfactory, with the possible exception of that specified by

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

with $b_1 = (y_1 + 2y_2 - y_3)/3$ and $b_2 = (y_1 - y_2 + 2y_3)/3$ having each the variance of $\frac{2}{3}$.

For $N = 3$, $p = 3$ the most efficient design is given by

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

with $b_1 = (y_1 + y_2 - y_3)/2$, and b_2 and b_3 given by cyclic permutation in this formula. The variance of each unknown is $\frac{3}{4}$.

For $N = 4$, $p = 3$ a design having an advantage in some situations is that given by

$$X = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

(together of course with those obtained by permutations of rows and of columns, as is to be understood throughout). The normal equations are

$$3b_1 + 2b_2 + 2b_3 = y_1 + y_2 + y_4$$

$$2b_1 + 3b_2 + 2b_3 = y_2 + y_3 + y_4$$

$$2b_1 + 2b_2 + 3b_3 = y_1 + y_3 + y_4.$$

An expeditious method of solution in this as in many similar cases is to add them all together and then subtract an appropriate multiple of the sum from each of the normal equations in turn. The variance of each unknown found by this

experiment is $5/7 = .714$. The simple experiment consisting of weighing one of the objects twice and the others once each yields variances in one case larger and in two cases smaller than this.

For $N = p = 4$ the cyclic arrangement

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

leads to variances all equal to $7/9$.

For $N = 5$, $p = 2$ the most efficient design appears to be

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

The variance of each unknown is in this case $1/3$.

REFERENCES

- [1] R. A. FISHER, *The Design of Experiments*, Oliver and Boyd, London and Edinburgh, 1935 and 1937.
- [2] F. YATES, "Complex experiments," *Jour. Roy. Stat. Soc. Supp.*, Vol. 2 (1935), pp. 181-247, including discussion. The reference is to p. 211.
- [3] R. A. FISHER, "The sampling error of estimated deviates. . .," *Mathematical Tables*, Vol. 1 (1931), pp. xxvi-xxxv. London, British Association for the Advancement of science.
- [4] P. C. TANG, "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat. Res. Mem.*, Vol. 2 (1939), pp. 126-149.
- [5] Cf. WHITTAKER and ROBINSON, *Calculus of Observations*, Section 115.
- [6] HAROLD HOTELLING, "Problems of prediction," *Am. Jour. of Sociology*, Vol. 48 (1942), pp. 61-76.

ON THE ANALYSIS OF A CERTAIN SIX-BY-SIX FOUR-GROUP LATTICE DESIGN¹

By BOYD HARSHBARGER²

Virginia Agricultural Experiment Station

1. Introduction. The lattice consists of groups of randomized incomplete blocks with certain restrictions being imposed on the randomization within each group, and the number of varieties is a perfect square. For example, if the number of varieties is $k^2 = 36$, then the orthogonal groups for a triple lattice, not considering randomizing within the blocks or between blocks, are as follows: (the numbers signify varieties).

GROUP X							GROUP Y						
Blocks							Blocks						
(1)	1	2	3	4	5	6	(1)	1	7	13	19	25	31
(2)	7	8	9	10	11	12	(2)	2	8	14	20	26	32
(3)	13	14	15	16	17	18	(3)	3	9	15	21	27	33
(4)	19	20	21	22	23	24	(4)	4	10	16	22	28	34
(5)	25	26	27	28	29	30	(5)	5	11	17	23	29	35
(6)	31	32	33	34	35	36	(6)	6	12	18	24	30	36

GROUP Z						
Blocks						
(1)	1	8	15	22	29	36
(2)	2	9	16	23	30	31
(3)	3	10	17	24	25	32
(4)	4	11	18	19	26	33
(5)	5	12	13	20	27	34
(6)	6	7	14	21	28	35

This design is constructed so that no variety appears with another variety more than once in the same block. This important characteristic makes the analysis simple, as it enables the results to be treated as a factorial design. The analysis is well described by Yates [3, 4, 5] and Cochran [1].

Suppose another group, U , is now formed from a six by six lattice, for example, the following group:

¹ Certain of the ideas presented here are embodied in the author's unpublished doctoral thesis by the same title, Library, George Washington University, Washington, D. C., 1943.

² The author wishes to express his appreciation to W. G. Cochran of Iowa State College, who advised freely in the preparation of the original thesis, and to Frank M. Weida of George Washington University.

GROUP U

Blocks						
(1)	31	26	21	16	11	6
(2)	1	32	27	22	17	12
(3)	7	2	33	28	23	18
(4)	13	8	3	34	29	24
(5)	19	14	9	4	35	30
(6)	25	20	15	10	5	36

The important characteristic, that no variety appears with another variety in the same block more than once, does not hold. For example, varieties 1 and 22 appear together in both groups Z and U.

It is the purpose of this paper to develop the statistical method for the analysis of such a design, where each group is duplicated, and to apply the results to an actual problem. The least square solution, as developed here, uses only the intra-block information to correct the varieties for block effects. In a second article the solution using both intra- and inter-block information will be given.

2. Estimation of the block and varietal effects. It is reasonable to assume in varietal trials that the general mean, and any particular block and variety effects, operate additively to produce the true mean of y associated with this block and variety. In particular, if y_{eij} is the yield of the plot for the j th variety in the i th block of the e th replicate, the following hypothesis may be set up, namely:

$$(1) \quad y_{eij} = \mu + \rho_e + \beta_{ei} + \nu_j + \epsilon_{eij}.$$

Where μ is the true or population mean yield, ρ_e is the population replicate effect of the e th replicate, β_{ei} is the population block effect of the i th block in the e th replicate, ν_j is the population variety effect of the j th variety, and ϵ_{eij} is the experimental error of the eij plot. Since the design has eight replicates, that is, each group is duplicated, the block effects may be estimated from unpaired and paired replicates or partners.

It is assumed that the ϵ_{eij} are independently and normally distributed with common variance $\frac{1}{W}$. Without loss of generality, it also may be assumed that the sum of the replicate effects, the sum of the block effects within any replication, and the sum of the variety effects are respectively equal to zero.

The parameters are estimated by the method of least squares, subject to the restrictions stated in the preceding paragraph. This involves choosing the parameters so that

$$(2) \quad S \left(y_{eij} - mx_1 - r_e x_{2e} - \frac{b_{ei} - b'_{ei}}{2} x_{3ei} - \frac{b_{ei} + b'_{ei}}{2} x_{4ei} - \nu_j x_{5j} \right)^2 \\ + \lambda_1 \sum_1^8 r_e + \sum_{e=1}^8 \lambda_{2e} \sum_{i=1}^6 \frac{b_{ei} - b'_{ei}}{2} + \sum_{e=1}^8 \lambda_{3e} \sum_{i=1}^6 \frac{b_{ei} + b'_{ei}}{2} + \lambda_4 \sum_{j=1}^{36} \nu_j$$

is a minimum.³ Here y_{eij} is the dependent variate, and x_1, \dots, x_6 , are the independent variates. In ordinary regression problems, the values of the x variates, as well as the y variate, constitute a part of the original data. However, in this case the y variate only is given, and the x variates must be constructed. Thus, for the design, one takes $x_1 = 1$ for all values; $x_2 = 1$ for all values in replicate e , but zero elsewhere; $x_{3ei} = 1$ for all values in the i th block of the e th replicate and -1 for all values in its partner, but zero elsewhere; $x_{4eij} = 1$ for all values in the i th block of the e th replicate and also 1 in its partner, but zero elsewhere, and $x_5 = 1$ where variety j occurs, but zero elsewhere.

One now takes the partial derivatives of the above equation with respect to the parameters and forms the normal equations. It can be shown that $(\lambda_1, \dots, \lambda_4)$ are each zero. The normal equations not involving λ 's are:

Leading term

$$\begin{aligned}
 (3) \quad m \quad & Nm + k^2 \sum_{e=1}^8 r_e + 2k \sum_{e=1}^8 \sum_{i=1}^6 \frac{b_{ei} - b'_{ei}}{2} \\
 & + 2k \sum_{e=1}^8 \sum_{i=1}^6 \frac{b_{ei} + b'_{ei}}{2} + 8 \sum_{j=1}^{36} v_j = G. \\
 r_e \quad & k_2 m + k^2 r_e + k \sum_{i=1}^6 b_{ei} + \sum_{j=1}^{36} v_j = R_e. \\
 \frac{b_{ei} - b'_{ei}}{2} \quad & k(r_e - r'_e) + 2k \frac{(b_{ei} - b'_{ei})}{2} = B_{ei} - B'_{ei}.
 \end{aligned}$$

Equations having $\frac{b_{ei} + b'_{ei}}{2}$ as leading terms.

Equations having v_j as leading terms.

In the above, N is the total number of values, k is the number of plots in a block, r_e is the e th replicate effect, b_{ei} is the i th block effect in the e th replicate, v_j is the j th variety effect, G is the total sum of all values, R_e is the sum of the values in the e th replicate, B_{ei} is the sum of values in the i th block of the e th replicate and v_j is the sum of the yields of the j th variety. The primes denote similar values of the partner terms.

Using the restrictions

$$(4) \quad \sum_{e=1}^8 r_e = \sum_{e=1}^8 \sum_{i=1}^6 \frac{b_{ei} + b'_{ei}}{2} = \sum_{e=1}^8 \sum_{i=1}^6 \frac{b_{ei} - b'_{ei}}{2} = \sum_{j=1}^{36} v_j = 0,$$

the values of the following parameters are directly obtainable:

$$(5) \quad m = \frac{G}{N}, \quad r_e = \frac{R_e}{k^2} - \frac{G}{N} \quad \text{and} \quad \frac{b_{ei} - b'_{ei}}{2} = \frac{B_{ei} - B'_{ei}}{2k} - \frac{r_e - r'_e}{2}.$$

³ Σ , for summation, will be used to represent summation over all values. Σ will be used in a more restricted sense.

The values of the $\frac{b_{si} + b'_{si}}{2}$ and v_i effects cannot be obtained directly. In order to simplify the solution of these equations, only the mean, confounded blocks, and the varietal effects will be used. The results later will be corrected for replicate effects. If each of the yields is added to the corresponding yield in its partner, one gets an equation of the form

$$(6) \quad y_{ik} = \mu + \beta_i + v_k + \epsilon_{ik}$$

where μ , β_i , and v_k are now twice their original values. These parameters are estimated by the method of least squares subject to the restrictions previously given. To distinguish them from the estimates derived in equation (3) they are designated with double primes (''). If B_{1x} is the total for block i in group X, that is from both pairs, and similarly for the other groups, the normal equations are:

Leading term

$$\begin{array}{lll} m'' & 144m'' & = G. \\ b''_{x1} & 6m'' + 6b''_{x1} + (v''_1 + v''_2 + v''_3 + v''_4 + v''_5 + v''_6) & = B_{1x}. \\ (7) \quad b''_{u6} & 6m'' + 6b''_{u6} + (v''_6 + v''_{10} + v''_{15} + v''_{20} + v''_{25} + v''_{30}) & = B_{u6}. \\ v''_1 & 4m'' + b''_{x1} + b''_{y1} + b''_{x1} + b''_{u2} + 4v''_1 & = V_1. \\ v''_6 & 4m'' + b''_{x6} + b''_{y6} + b''_{x1} + b''_{u3} + 4v''_6 & = V_{36}. \end{array}$$

Let T_{xi} be the total of all the varieties appearing in block i of group X from all the replicates; T_{yi} the total of all values appearing in block i of group Y from all the replicates, etc. Also

$$(8) \quad C_{xi} = 4B_{xi} - T_{xi} \quad C_{yi} = 4B_{yi} - T_{yi} \quad \text{etc.}$$

Solving the equations:

$$b''_{xi} = \frac{C_{xi}}{18} \quad b''_{yi} = \frac{C_{yi}}{18}$$

and

$$\begin{array}{l} b''_{x1} = \frac{1}{48}[(25C_{x1} + C_{x3} + C_{x5}) + 3(C_{u2} + C_{u4} + C_{u6})] \\ (9) \quad b''_{x2} = \frac{1}{48}[(25C_{x2} + C_{x4} + C_{x6}) + 3(C_{u1} + C_{u3} + C_{u5})] \\ b''_{u6} = \frac{1}{48}[(25C_{u6} + C_{u2} + C_{u4}) + 3(C_{x1} + C_{x3} + C_{x5})]. \end{array}$$

The values of b'' 's calculated as above contain the replicate effects. To correct for this, adjust the values so that the sum of the block effects for each replicate is zero. From the corrected b'' values and the normal equations with the v'' 's as leading terms, the corrected varietal sums are calculated. These are:

$$(10) \quad 4v''_j + 4m'' = V_j - \Sigma b''_{ji} \quad (\text{sum over blocks in which } v_j \text{ appears})$$

where $4v''_j + 4m''$ is the corrected varietal total for the j th variety.

3. Test of significance and the analysis of variance. If the x 's have the previously defined values, the following identity occurs:

$$\begin{aligned}
 (11) \quad Sy^2 = & mG + \sum_{s=1}^8 r_s R_s + \sum_{s=1}^8 \sum_{i=1}^6 \frac{(b_{si} - b'_{si})}{2} (B_{si} - B'_{si}) \\
 & + \sum_{s=1}^8 \sum_{i=1}^6 \frac{(b_{si} + b'_{si})}{2} (B_{si} + B'_{si}) + \sum_{j=1}^{36} v_j V_j \\
 & + S \left(y - mx_1 - r_s x_{2s} - \frac{b_{si} - b'_{si}}{2} x_{3i} - \frac{b_{si} + b'_{si}}{2} x_{4i} - v_j x_{5j} \right)^2.
 \end{aligned}$$

In the equation (11)

$$\begin{aligned}
 (12) \quad mG + \sum_{s=1}^8 r_s R_s + \sum_{s=1}^8 \sum_{i=1}^6 \frac{(b_{si} - b'_{si})}{2} (B_{si} - B'_{si}) \\
 + \sum_{s=1}^8 \sum_{i=1}^6 \frac{(b_{si} + b'_{si})}{2} (B_{si} + B'_{si}) + \sum_{j=1}^{36} v_j V_j,
 \end{aligned}$$

is the reduction in the sum of squares due to regression and

$$(13) \quad S \left(y - mx_1 - r_s x_{2s} - \frac{b_{si} - b'_{si}}{2} x_{3i} - \frac{(b_{si} + b'_{si})}{2} x_{4i} - v_j x_{5j} \right)^2,$$

is the residual sum of squares. The reductions attributable to

$$\sum_{s=1}^8 \sum_{i=1}^6 \frac{(b_{si} - b'_{si})}{2} (B_{si} - B'_{si})$$

and

$$(14) \quad \sum_{s=1}^8 \sum_{i=1}^6 \frac{(b_{si} + b'_{si})}{2} (B_{si} + B'_{si}),$$

will be designated as component (a) and component (b) respectively. The residual mean square will be denoted by s^2 .

The common test required is that of the null hypothesis that the variety effects v_1, v_2, \dots, v_{36} are all zero. This test is made by calculating the reduction (R_t) to the sum of squares on all variates, and the reduction (R_v) due to the regression on all variates, except the variety effects. $R_t - R_v$ is called the additional reduction to the sum of squares due to the v 's after fitting the remaining variates.

The ratio $(R_t - R_v)/(82 - 47)s^2$ is distributed as F , as shown by Yates (6), with 35 and 205 degrees of freedom. The 35, 205, 82, and 47 degrees of freedom pertain respectively to varieties, error, all constants fitted, and the total constants less the constants for varieties.

Referring to formula (11) and the parameter effects, the sum of squares in the "Analysis of Variance Table" follow directly for replicate and component (a). Nair² in his recent article gives in detail the method for getting out the reduction

to the sum of squares for the entangled components. He shows that the reductions for component (b) and the varieties may be written as:

$$(15) \quad \frac{1}{8} \sum_{e=x}^u \sum_{i=1}^6 b''_{ei} C_{ei} \quad \text{and} \quad \sum_{j=1}^{36} \frac{V_j^2}{8} - \frac{G^2}{N}$$

where the b'' have been corrected for replicate effects. It is well to note that $\frac{1}{8} \sum_{e=x}^u \sum_{i=1}^6 b''_{ei} C_{ei}$ is the reduction due to intra-block effects, freed of varietal effects.

The reduction due to varieties corrected for block effects is given by

$$(16) \quad \frac{1}{8} \sum_{e=x}^u \sum_{i=1}^6 b''_{ei} C_{ei} + \left(\sum_{j=1}^{36} \frac{V_j^2}{8} - \frac{G^2}{N} \right) - \left(\sum_{e=x}^u \sum_{i=1}^6 \frac{B_{ei}^2}{6} - \frac{G^2}{N} \right).$$

This reduction can be used for testing the variety effects.

ANALYSIS OF VARIANCE TABLE

	D/F	
<i>Replicate</i>	(8-1)	$\sum_{e=1}^8 \frac{R_e^2}{36} - \frac{G^2}{N}$
<i>Component (a)</i>	4(6-1)	$\frac{1}{8} \sum_{e=1}^8 \sum_{i=1}^6 \frac{(B_{ei} - B'_{ei})^2}{12} - \sum_{i=1}^6 \frac{(R_i - R'_i)^2}{72}$
<i>Component (b)</i>	4(6-1)	$\frac{1}{8} \sum_{e=x}^u \sum_{i=1}^6 b''_{ei} C_{ei}$
<i>Varieties</i>		
(ignoring blocks)	(36-1)	$\sum_{j=1}^{36} \frac{V_j^2}{8} - \frac{G^2}{N}$
<i>Error</i>	205	obtained by subtraction
<i>Total</i>	8(36)-1	$Sy^2 - \frac{G^2}{N}$

4. Standard error of adjusted varietal means. For obtaining the standard error of the difference between two varieties adjusted by the intra-block information, this difference between two varieties can be expressed as a linear function of the plot yields. The standard error of the difference then can be obtained from the standard error of a linear function. To obtain the coefficients, it is well to draw a sketch of the plots, and put the coefficient of each plot on the diagram. In this way the proper multipliers can be found in a convenient manner.

First consider the case where the two varieties appear together in the same block in both groups Z and U. One such pair consists of varieties (1) and (22), for which the varietal effects are designated by v''_1 and v''_2 . From equation (10) we have:

$$(17) \quad \begin{aligned} 4v''_1 &= V_1 - 4m'' - b''_{x1} - b''_{y1} - b''_{z1} - b''_{u1} \\ 4v''_2 &= V_{22} - 4m'' - b''_{x4} - b''_{y4} - b''_{z1} - b''_{u2} \end{aligned}$$

The linear function of the difference between the varietal effects is:

$$(18) \quad 4(v''_{22} - v''_1) = V_{22} - V_1 + b''_{x1} - b''_{x4} + b''_{y1} - b''_{y4}$$

where

$$b''_{x1} = \frac{1}{3k} (4B_{x1} - T_{x1}) \quad b''_{x4} = \frac{1}{3k} (4B_{x4} - T_{x4})$$

$$b''_{y1} = \frac{1}{3k} (4B_{y1} - T_{y1}) \quad b''_{y4} = \frac{1}{3k} (4B_{y4} - T_{y4}).$$

The multipliers [except for the common factor 4 shown on the left of equation (18)] are:

<i>Number of Plots</i>	<i>Multipliers</i>	<i>Contribution to variance</i>
4	$\pm \frac{3k+2}{3k}$	$\frac{36k^2 + 48k + 16}{W(9k^2)}$
4	$\pm \frac{3k-2}{3k}$	$\frac{36k^2 - 48k + 16}{W(9k^2)}$
4	$\pm \frac{4}{3k}$	$\frac{64}{W(9k^2)}$
$4(k+2)$	$\pm \frac{3}{3k}$	$\frac{36(k-2)}{W(9k^2)}$
$12(k+2)$	$\pm \frac{1}{3k}$	$\frac{12(k-2)}{W(9k^2)}$
Total		$\frac{72k^2 + 48k}{W(9k^2)}$

The variance per plot of the difference between two varietal means for varieties which occur together in the same block in groups Z and U is:

$$(19) \quad \frac{72k^2 + 48k}{2W(16)(9k)} = \frac{3k+2}{12Wk}$$

and for $k = 6$ is $\frac{5}{18W}$.

Similarly the variance per plot of the difference between varietal means for other combinations are as follows:

<i>Combinations</i>	<i>formula for $k = 6$</i>
<i>both in same block in groups Z or U</i>	$\frac{7}{24W}$
<i>both in same block in groups X or Y</i>	$\frac{8}{27W}$
<i>not together in the same block</i>	$\frac{67}{216W}$

5. Numerical Analysis. (a) The data. In order to illustrate the application of the method developed, an experiment used to test the yields of 36 hybrid corn varieties is presented. This experiment was carried out on the Arlington Experimental Farm, and the results are used through the courtesy of A. E. Brandt⁴ and M. H. Jenkins.⁵

Except for randomization, the plot yields are as shown in tables I to IV.

(b). *Calculation for analysis of variance table.* From page 9, the total sum of squares, the sum of squares for replicates, and the sum of squares for varieties ignoring blocks are obtained by substitution. They are:

$$\begin{aligned} (90.9)^2 + (81.4)^2 + \dots + (101.0)^2 - c &= 33,546.92 \\ \frac{(3291.8)^2 + (3300.3)^2 + \dots + (2978.2)^2}{36} - c &= 2,289.68 \end{aligned}$$

and

$$\frac{(741.2)^2 + (695.6)^2 + \dots + (743.1)^2}{8} - c = 15,825.09$$

where

$$c = \frac{(25,935.9)^2}{288} = 2,335,662.80$$

The block sum of squares, eliminating varieties, is made up of two parts, component (a) and component (b). From the formula on page 9 the reduction for component (a) is:

$$\begin{aligned} \frac{(559.2 - 540.2)^2 + (547.0 - 522.4)^2 + \dots + (515.8 - 507.7)^2}{12} \\ - \frac{(3291.8 - 3300.2)^2 + (3256.5 - 3304.7)^2 + \dots + (3284.6 - 2978.2)^2}{72} = 1,415.27. \end{aligned}$$

Component (b) consists of differences giving an estimate of block yield freed of varietal effects. The C 's are first calculated by using formula (8) and the results are as follows:

$$C_{x1} = (4B_{x1} - T_{x1}) = 4(1,099.4) - 4203.2 = 194.4 \text{ etc.}$$

The b 's are calculated by using formulas given by (9), and then correcting for replicate effects by imposing the conditions that

$$\sum_{i=1}^6 b''_{ei} = 0 \quad (e = x \dots u)$$

The corrected b''_{ei} are:

$$b''_{11} = 6.79556, \quad b''_{21} = -3.83777, \quad \dots \quad b''_{24} = -0.34306.$$

⁴ Acting Chief, Conservation Experiment Station Divisions, Office of Research Soil Conservation Service, U.S.D.A.

⁵ Principal Agronomist In charge of Corn Investigations, Bureau of Plant Industry, U.S.D.A.

TABLE I
GROUP X
Plot Yield Replicate 1

						Σ
1	2	3	4	5	6	
90.9	81.4	95.4	96.9	96.6	98.0	559.2
7	8	9	10	11	12	
96.0	92.9	86.3	96.4	84.5	90.9	547.0
13	14	15	16	17	18	
79.3	97.5	95.9	80.2	101.8	112.1	566.8
19	20	21	22	23	24	
83.4	92.6	91.2	88.3	71.1	111.8	538.4
25	26	27	28	29	30	
93.7	86.1	68.0	109.2	94.9	93.3	545.2
31	32	33	34	35	36	
85.3	81.2	99.5	86.9	93.4	88.9	535.2
Σ	528.6	531.7	536.3	557.9	542.3	595.0
						3291.8

Plot Yield Replicate 1'

						Σ
1	2	3	4	5	6	
105.8	83.5	68.6	99.0	91.8	91.5	540.2
7	8	9	10	11	12	
98.6	71.9	81.0	98.6	90.3	82.0	522.4
13	14	15	16	17	18	
70.5	90.4	86.2	88.8	99.7	113.5	549.1
19	20	21	22	23	24	
94.6	89.6	106.8	99.2	73.2	90.9	554.3
25	26	27	28	29	30	
91.6	86.9	74.6	104.1	102.3	96.0	555.5
31	32	33	34	35	36	
91.1	95.1	98.8	95.5	88.1	110.2	578.8
Σ	552.2	516.4	516.0	585.2	545.4	584.1
						3300.3

TABLE II
GROUP Y
Plot Yield Replicate 2

						Σ
1 85.9	7 96.6	13 85.4	19 86.4	25 88.8	31 89.7	532.8
2 87.1	8 83.5	14 86.9	20 88.1	26 89.1	32 102.7	537.4
3 88.4	9 85.5	15 90.4	21 99.7	27 80.5	33 100.2	544.7
4 88.5	10 72.3	16 89.5	22 95.6	28 105.5	34 82.6	534.0
5 87.0	11 88.1	17 101.8	23 82.5	29 86.8	35 100.1	546.3
6 90.8	12 83.4	18 101.2	24 120.2	30 72.6	36 93.1	561.3
						3256.5

Plot Yield Replicate 2'

						Σ
1 95.1	7 96.0	13 85.4	19 74.3	25 93.0	31 86.7	530.5
2 96.5	8 84.2	14 79.0	20 88.1	26 95.6	32 101.3	544.7
3 95.4	9 83.3	15 95.9	21 99.0	27 66.0	33 105.3	544.9
4 81.4	10 95.0	16 91.6	22 101.4	28 108.5	34 90.5	568.4
5 91.8	11 107.6	17 84.2	23 66.1	29 87.6	35 101.6	538.9
6 92.2	12 89.6	18 104.1	24 116.0	30 81.5	36 93.9	577.3
						3304.7

TABLE III
GROUP Z
Plot Yield Replicate 3

						Σ
1 87.3	8 79.9	15 86.9	22 105.0	29 99.4	36 90.3	548.8
2 90.0	9 93.0	16 93.1	23 90.3	30 73.9	31 94.1	534.4
3 91.2	10 92.9	17 90.6	24 92.3	25 76.0	32 97.2	540.2
4 87.8	11 98.2	18 125.1	19 102.1	26 88.3	33 98.1	599.6
5 84.3	12 93.7	13 73.9	20 84.4	27 66.7	34 89.8	492.8
6 90.1	7 91.3	14 85.4	21 95.4	28 93.7	35 101.6	557.5
						3273.3

Plot Yield Replicate 3'

						Σ
1 99.4	8 83.5	15 70.9	22 99.9	29 103.0	36 69.0	525.7
2 92.2	9 96.0	16 98.1	23 74.7	30 89.8	31 103.6	554.4
3 91.9	10 106.4	17 92.7	24 108.3	25 88.8	32 103.4	591.5
4 99.0	11 105.4	18 101.2	19 78.1	26 90.5	33 98.1	572.3
5 63.7	12 94.4	13 80.7	20 78.4	27 71.3	34 86.9	475.4
6 81.5	7 88.6	14 90.4	21 84.1	28 93.7	35 88.9	527.2
						3246.5

TABLE IV
GROUP U
Plot Yield Replicate 4

						Σ
31 98.5	26 99.3	21 97.6	16 90.9	11 102.5	6 91.5	580.3
1 95.9	32 92.3	27 70.0	22 99.2	17 85.6	12 100.6	543.6
7 77.8	2 79.9	33 98.1	28 104.8	23 88.9	18 115.7	565.2
13 78.6	8 81.3	3 84.1	34 86.9	29 91.3	24 111.8	534.0
19 82.6	14 89.0	9 75.0	4 94.1	35 108.3	30 96.7	545.7
25 87.3	20 76.2	15 89.7	10 83.7	5 82.2	36 96.7	515.8
						3284.6

Plot Yield Replicate 4'

						Σ
31 83.1	26 88.3	21 84.8	16 94.5	11 73.6	6 84.4	508.7
1 80.9	32 82.6	27 73.9	22 94.8	17 75.8	12 83.4	491.4
7 93.3	2 85.0	33 100.2	28 87.8	23 83.2	18 94.7	544.2
13 74.6	8 68.2	3 45.2	34 79.0	29 87.6	24 102.1	456.7
19 69.1	14 81.2	9 80.3	4 68.8	35 83.7	30 86.4	469.5
25 89.5	20 71.7	15 75.1	10 97.1	5 73.3	36 101.0	507.7
						2978.2

To get the reduction due to component (b) the above results are substituted in

$$\frac{1}{8} \sum_{e=2}^u \sum_{i=1}^6 b''_{ei} C_{ei} = 1,389.96.$$

The necessary results are now available for the "Analysis of Variance Table."

THE ANALYSIS OF VARIANCE TABLE

<i>Source of Variation</i>	<i>Degrees of Freedom</i>	<i>Sum of Squares</i>	<i>Mean Square</i>
<i>Replications</i>	7	2,289.68	327.097
<i>Component (a)</i>	20	1,415.27	
<i>Component (b)</i>	20	1,389.96	
<i>Blocks (eliminating varieties)</i>	40	2,805.23	70.131
<i>Varieties (ignoring blocks)</i>	35	15,825.09	
<i>Error</i>	205	12,626.92	61.595
<i>Total</i>	287	33,546.92	

(c) *Test of significance* There frequently will be large differences between varieties so that a test of significance may not be needed. If a test is needed, one involving only intra-block information may be used. For this purpose, it is necessary to calculate the sum of squares for varieties eliminating block effects as shown by formula (15): 13,946.28. The mean square will be 399.893, and

$$F = \frac{399.893}{61.595} = 6.49 \text{ which is highly significant.}$$

(d). *Corrected varietal totals and means.* The right-hand side of equation (10) gives the corrected variety totals, and when divided by eight gives the varietal means. These corrected varietal means can then be compared to determine the best variety. The corrected varietal totals and means are:

Corrected Varietal Totals

1 743.22	2 669.05	3 652.00	4 705.80	5 672.04	6 720.32
7 747.59	8 658.58	9 664.57	10 751.39	11 739.54	12 735.95
13 642.31	14 700.44	15 686.41	16 713.21	17 730.79	18 857.26
19 665.95	20 675.40	21 756.25	22 801.34	23 619.46	24 868.84
25 704.17	26 699.83	27 567.71	28 814.04	29 763.51	30 679.44
31 721.79	32 757.48	33 783.42	34 726.05	35 780.48	36 760.37

Corrected Varietal Means

1 92.902	2 83.631	3 81.500	4 88.225	5 84.005	6 90.040
7 93.449	8 82.322	9 83.071	10 93.924	11 92.442	12 91.994
13 80.289	14 87.555	15 85.801	16 89.151	17 91.349	18 107.158
19 83.244	20 84.425	21 94.531	22 100.168	23 77.432	24 108.605
25 88.021	26 87.479	27 70.964	28 101.755	29 95.439	30 84.930
31 90.224	32 94.685	33 97.927	34 90.756	35 97.560	36 93.046

When comparing one variety with another it is necessary to know the standard error of the mean difference, in order to judge whether this difference is significant. The formulas for the standard error of the difference between mean yields differ for those sets of varieties which occur together in the same block in groups Z and U, in groups Z or U, in groups X or Y, or do not occur together in the same block. The formulas for these calculations are, respectively:

(19), (20), (21), and (22). For example, the standard error of the difference between two variety means, such as variety 1 and 2, is

$$\frac{8}{27}(61.595) = 4.27.$$

REFERENCES

- [1] G. M. COX, R. C. ECKHARDT and W. G. COCHRAN, "The analysis of lattice and triple lattice experiments in corn varietal tests," *Iowa Agr. Exp. Sta. Res. Bul.*, Vol. 281 (1940).
- [2] K. P. NAIR, "A note on the method of fitting constants," *Sankhya*, Vol. 5 (1941), pp. 317-328.
- [3] F. YATES, "The principles of orthogonality and confounding in replicated experiments," *Journal Agr. Sci.*, Vol. 23 (1933), 108-145.
- [4] F. YATES, "Complex experiments," *Suppl. Jour. Roy. Stat. Soc.*, Vol. 2 (1935), pp. 181-247.
- [5] F. YATES, "A new method of arranging variety trials involving a large number of varieties," *Journal Agr. Sci.*, Vol. 26 (1936), pp. 424-465.
- [6] F. YATES, "Orthogonal functions and tests of significance in the analysis of variance," *Journal Royal Stat. Soc.*, Vol. 5 (1938), pp. 177-180.

NOTES

This section is devoted to brief research and expository articles on methodology and other short items.

ON THE EXPECTED VALUES OF TWO STATISTICS

BY H. E. ROBBINS

Post Graduate School; Annapolis, Md.

In a previous paper¹, the following theorem was proved. Let X be a random, Lebesgue measurable subset of Euclidean m dimensional space E_m , and let $\mu(X)$ be the measure of X . For every point x of E_m let $p(x)$ be the probability that X contains x . Then

$$(1) \quad E(\mu(X)) = \int_{E_m} p(x) d\mu(x).$$

In the present note we shall show how this theorem may be used to find the expected values of two statistics which arise in sampling theory. Applications to similar problems may suggest themselves to the reader.

Let Y be a real random variable with c. d. f. (cumulative distribution function) $\sigma(y)$, so that for every y ,

$$(2) \quad \Pr(Y < y) = \sigma(y).$$

Let Y_1, \dots, Y_n be n independent random variables, each with the distribution of Y . Finally, let

$$(3) \quad \begin{aligned} A &= \min(Y_1, \dots, Y_n), \\ B &= \max(Y_1, \dots, Y_n), \\ R &= B - A, \\ F &= \sigma(B) - \sigma(A). \end{aligned}$$

Although the values of $E(F)$ and $E(R)$ can be found from the sampling distributions of F and R , and, in fact, are well known, we shall show how to apply (1) to find $E(F)$ and $E(R)$ directly.

To find the first of these, let X denote the set of points in the interval $0 \leq x \leq 1$ such that

$$(4) \quad \sigma(A) < x < \sigma(B).$$

Then X is a random set with measure

$$(5) \quad \mu(X) = F.$$

Moreover, for any point x the probability that X shall contain x is clearly

$$(6) \quad p(x) = 1 - x^n - (1 - x)^n.$$

Hence by (1),

$$(7) \quad E(\mu(X)) = \int_0^1 [1 - x^n - (1 - x)^n] dx = \frac{n-1}{n+1}.$$

Thus by (5),

$$(8) \quad E(F) = \frac{n-1}{n+1}.$$

This result may also be derived by the usual method. In fact, it is not hard to show that the c. d. f. of F is

$$(9) \quad \tau(f) = \Pr(F < f) = (1-n)f^n + nf^{n+1} \quad \text{for } 0 \leq f \leq 1,$$

whence

$$(10) \quad \begin{aligned} E(F) &= \int_0^1 f d\tau(f) = (1-n)n \int_0^1 f^n df + n(n-1) \int_0^1 f^{n+1} df \\ &= \frac{(1-n)n}{n+1} + \frac{n(n-1)}{n} = \frac{n-1}{n+1}. \end{aligned}$$

Here the advantage of using (1) is only that it makes unnecessary the calculation of the c. d. f. $\tau(f)$, provided that only $E(F)$ is desired.

The situation is quite otherwise with $E(R)$. Here the c. d. f. of R is

$$(11) \quad \theta(r) = \Pr(R < r) = n(n-1) \int_{-\infty}^{\infty} \varphi(a) \int_a^{a+r} \varphi(b) \left[\int_a^b \varphi(t) dt \right]^{n-2} db da,$$

where φ is the probability density function of Y . Unless φ is a very simple function, it would seem difficult to find a simple expression for the integral

$$(12) \quad E(R) = \int_0^{\infty} r d\theta(r).$$

However, if we let X now denote the linear set

$$(13) \quad A \leq t \leq B,$$

then

$$(14) \quad \mu(X) = B - A = R.$$

The probability that X shall contain the point t is now

$$(15) \quad p(t) = 1 - \sigma^n(t) - (1 - \sigma(t))^n,$$

so that, by (1) and (14),

$$(16) \quad E(R) = \int_{-\infty}^{\infty} \{1 - \sigma^n(t) - (1 - \sigma(t))^n\} dt.$$

This formula for the expected value of the range in a sample of n from a population Y with c. d. f. $\sigma(t)$ is believed to be new.

If $\sigma(t)$ is such that $dt/d\sigma$ can be found as an explicit function of σ , then (16) can be written with advantage as

$$(17) \quad E(R) = \int_0^1 \{1 - \sigma^n - (1 - \sigma)^n\} \frac{dt}{d\sigma} d\sigma.$$

For example, suppose the random variable Y has the probability density function

$$(18) \quad \varphi(y) = \frac{e^y}{(1 + e^y)^2},$$

and hence the c. d. f.

$$(19) \quad \sigma(y) = \frac{e^y}{1 + e^y}.$$

Then

$$(20) \quad t = \log \frac{\sigma}{1 - \sigma}, \quad \frac{dt}{d\sigma} = \frac{1}{\sigma(1 - \sigma)}.$$

Hence from (17), the expected value of the range in a sample of n is

$$(21) \quad E(R) = \int_0^1 \frac{1 - \sigma^n - (1 - \sigma)^n}{\sigma(1 - \sigma)} d\sigma.$$

The indicated division in the integrand may be carried out, and the result, a polynomial in σ of degree $\leq (n - 2)$, when integrated between 0 and 1, gives an explicit formula for $E(R)$. Thus for samples of $n = 2, 3, 4$ we find the values of $E(R)$ to be respectively 2, 3, 11/3. Incidentally, it is always true that the expected value of the range for $n = 3$ is three-halves that for $n = 2$. This follows from (16) and the algebraic identity

$$(22) \quad \{1 - \sigma^3 - (1 - \sigma)^3\} = \frac{3}{2}\{1 - \sigma^2 - (1 - \sigma)^2\}.$$

REFERENCES

- [1] H. E. ROBBINS, "On the measure of a random set," *Annals of Math Stat*, Vol 15 (1944), pp. 70-74.

ON RELATIVE ERRORS IN SYSTEMS OF LINEAR EQUATIONS

BY A. T. LONSETH

Northwestern University

Some time ago in these *Annals*¹, L. B. Tuckerman discussed the effect of relative coefficient errors on relative solution errors for a non-singular linear algebraic system; his discussion was confined to errors so small that their squares and higher powers can be neglected. Dr Tuckerman's paper was principally concerned

¹ L. B. Tuckerman, "On the mathematically significant figures in the solution of simultaneous linear equations," *Annals of Math. Stat*, Vol. 12 (1941), pp. 307-316.

with the important problem of limiting errors incurred while solving the system, and has suggested the desirability of a non-infinitesimal treatment of relative errors. Such a treatment follows; the method is a variant of that used in a paper on absolute errors². The computations provide (1) a criterion for allowable relative errors in the coefficients to assure non-vanishing of the determinant; (2) a bound (subject to this criterion) for the relative error in each solution; (3) a specification of accuracy in the coefficients to assure a pre-assigned accuracy in the solution.

We consider a system of linear equations

$$(1) \quad \sum_{i=1}^n a_{ij} x_i = c_j \quad j = 1, 2, \dots, n,$$

where none of the following vanish: the $n(n+1)$ coefficients a_{ij} , c_i ; the determinant Δ of the system; and the n solution-components x_i . Under these conditions it is possible to speak of "relative errors" in the a 's, c 's, and x 's. Let ϵ_{ij} , σ_i be the relative errors in a_{ij} and c_i respectively, so that a_{ij} is perturbed to $a_{ij}(1 + \epsilon_{ij})$; c_i to $c_i(1 + \sigma_i)$. We inquire as to limitations on ϵ_{ij} and σ_i which will permit solution for $x_j(1 + \rho_j)$ of the system

$$(2) \quad \sum_{i=1}^n a_{ij}(1 + \epsilon_{ij})x_j(1 + \rho_j) = c_i(1 + \sigma_i) \quad i = 1, 2, \dots, n,$$

where ρ_j is the relative error induced in x_j ; and we seek to limit $|\rho_j|$ in terms of the ϵ 's and σ 's. We shall assume that for all i, j

$$(3) \quad |\epsilon_{ij}|, |\sigma_i| < \delta,$$

where δ will be suitably restricted later.

Combining (1) and (2) we get

$$(4) \quad \sum_{j=1}^n a_{ij} x_j \rho_j = c_i \sigma_i - \sum_{j=1}^n a_{ij} \epsilon_{ij} x_j - \sum_{j=1}^n a_{ij} \epsilon_{ij} x_j \rho_j \quad i = 1, 2, \dots, n,$$

Since by hypothesis the determinant Δ of (1) is not zero, matrix $A = (a_{ij})$ has the inverse $A^{-1} = (b_{ji}) = (A_{ji}/\Delta)$, where A_{ji} is the cofactor of a_{ij} in Δ . Multiplying (4) by b_{ki} and summing on i we get

$$(5) \quad x_k \rho_k = \sum_{i=1}^n b_{ki} c_i \sigma_i - \sum_{i=1}^n b_{ki} \sum_{j=1}^n a_{ij} \epsilon_{ij} x_j \\ - \sum_{i=1}^n b_{ki} \sum_{j=1}^n a_{ij} \epsilon_{ij} x_j \rho_j \quad k = 1, 2, \dots, n,$$

and by (3)

$$(6) \quad \rho \leq \delta(M_k + N_k |\rho_k|),$$

² A. T. Lonseth, "Systems of linear equations with coefficients subject to error," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 332-337.

where ρ is the greatest $|\rho_k|$, and

$$(7) \quad M_k = \frac{1}{|x_k|} \sum_{i=1}^n |b_{ki}| \left(|c_i| + \sum_{j=1}^n |a_{ij} x_j| \right),$$

$$(8) \quad N_k = \frac{1}{|x_k|} \sum_{i=1}^n \sum_{j=1}^n |b_{ki} a_{ij} x_j|,$$

so that

$$M_k = N_k + \frac{1}{|x_k|} \sum_{i=1}^n |b_{ki} c_i|.$$

Denote by M, N the maximum values of M_k, N_k respectively over $k = 1, 2, \dots, n$. From (6),

$$\rho \leq \delta(M + N\rho),$$

whence, if

$$(9) \quad \delta < 1/N,$$

it follows that

$$(10) \quad \rho \leq \delta M / (1 - \delta N),$$

which of course bounds each individual $|\rho_k|$, though rather crudely. To bound $|\rho_k|$ more genuinely in terms of δ, M_k, N_k, M and N it remains only to use this inequality with (6):

$$(11) \quad |\rho_k| \leq \delta(M_k + \delta M N_k / (1 - \delta N)), \quad k = 1, 2, \dots, n;$$

with M_k, N_k as given in (7) and (8).

If (9) holds, then, it follows that $|\rho_k|$ is bounded by (11)—if ρ_k exists. This essential point can be established by solving (5) for ρ_k by iteration²; (9) is a sufficient condition for convergence of the resulting series, and hence for non-singularity of the perturbed matrix $(a_{ij} + \epsilon_{ij} a_{ij})$.

In order to be sure that $|\rho_k| \leq \eta$, a pre-assigned number, for all k , it suffices by (10) to choose δ so that

$$\delta M / (1 - \delta N) \leq \eta,$$

whence

$$\delta \leq \eta / (M + N\eta).$$

A less simple inequality whose satisfaction by δ will guarantee that $|\rho_k| \leq \eta_k$ follows from (11), namely

$$\delta \leq (A - B) / 2C,$$

where $A = \{(M_k - N\eta_k)^2 + 4MN_k\eta_k\}^{\frac{1}{2}}$, $B = M_k + N\eta_k$, $C = |MN_k - NM_k|$.

A RECIPROCITY PRINCIPLE FOR THE NEYMAN-PEARSON THEORY OF TESTING STATISTICAL HYPOTHESES

BY LOUIS M. COURT

In contrasting the tested hypothesis H_1 with the alternative H_2 , i.e., in comparing the probability distribution $p(x_1, \dots, x_n | H_1)$ associated with the first hypothesis with the distribution $p(x_1, \dots, x_n | H_2)$ associated with the second, Neyman and Pearson select the *best* critical region R^* from the infinite set of critical regions R of a specified size α by minimizing the probability

$$(1) \quad \int_{S-R} p(x_1, \dots, x_n | H_2) dx_1 \cdots dx_n$$

of accepting H_1 when H_2 is true (Type II Errors) subject to the constancy of the probability of rejecting H_1 when H_1 is correct (Type I Errors),

$$(2) \quad \int_R p(x_1, \dots, x_n | H_1) dx_1 \cdots dx_n = \alpha.$$

S in (1) represents the whole of variate (x_1, \dots, x_n) space and $S - R$, the complement of R relative to S .

Obviously (1) is conditionally minimized when

$$(3) \quad \int_R p(x_1, \dots, x_n | H_2) dx_1 \cdots dx_n$$

is maximized subject to (2). Neyman and Pearson have shown that if one or more members of the one parameter (λ) family of regions $R^*(\lambda)$ defined by the inequalities

$$(4) \quad p(x_1, \dots, x_n | H_2) \geq \lambda p(x_1, \dots, x_n | H_1)$$

satisfy the "side" condition (2), they will be best critical regions maximizing (3) subject to (2) or minimizing (1) subject to (2)¹. As suggested by the notation, the family $R^*(\lambda)$ depends upon λ and, if sufficient restrictions are imposed upon $p(x_1, \dots, x_n | H_1)$ and $p(x_1, \dots, x_n | H_2)$, there is *one and only one* region for every value of λ lying in an interval contained in the positive half-axis. λ , itself, is clearly a function $\lambda(\alpha)$ of α . Consequently, $R^*(\lambda)$ depends upon α and may be written as $R^*(\alpha)$. The best critical region for a preassigned size $\bar{\alpha}$ is given by $R^*(\bar{\alpha})$.

Will we get the *same* best critical region if among the regions T that fix the probability of Type II Errors,

$$(5) \quad \int_{S-T} p(x_1, \dots, x_n | H_2) dx_1 \cdots dx_n = 1 - \beta,$$

¹ For a full exposition, see Neyman and Pearson, *Stat. Res. Memoirs*, Vol. 1 (1936)

we find the one that minimizes the integral in (2) with R replaced by T , i.e. if we find the one that minimizes the probability of Type I Errors? We shall call this turnabout of the usual process the *reversed* Neyman-Pearson principle.

To discover the answer, we note that $\int_S p(x_1, \dots, x_n | H_2) dx_1 \dots dx_n$ is equal to unity and (5) may be rewritten as

$$(6) \quad \int_T p(x_1, \dots, x_n | H_2) dx_1 \dots dx_n = \beta.$$

The regions that minimize the left side of (2) with R replaced by T subject to (6) are obviously identical with those that maximize the negative of this left side subject to (6) multiplied through by -1 . The latter problem is formally identical with the one referred to in the second paragraph of this note and, invoking Neyman and Pearson's result, we conclude that the said conditional maximization is effected by the one parameter (μ) family of regions $T^*(\mu^{-1})$ defined by the inequalities

$$(7) \quad -p(x_1, \dots, x_n | H_1) \geq -\mu p(x_1, \dots, x_n | H_2)$$

$$\text{or} \quad p(x_1, \dots, x_n | H_2) \geq \frac{1}{\mu} p(x_1, \dots, x_n | H_1).$$

μ^{-1} in $T^*(\mu^{-1})$ denotes the reciprocal of μ . It is clear from (4) and (7) that the families of regions $R^*(\lambda)$ and $T^*(\mu^{-1})$, satisfying the direct and reversed Neyman-Pearson processes, coincide

As before, μ is some function $\mu(\beta)$ of β . Hence, β is a function $\beta(\mu)^2$ of μ and, accordingly, a function $\beta \left[\frac{1}{\lambda(\alpha)} \right]$ of α . Consequently, if the level at which the probability of Type II Errors in the *reversed* Neyman-Pearson process is held constant is taken equal to $1 - \beta \left[\frac{1}{\lambda(\alpha)} \right]$ in terms of the level α at which the probability of Type I Errors is fixed in the *direct* Neyman-Pearson method, the reversed and direct processes yield the *same* best critical region. *This is the reciprocity principle alluded to in the title of this note in its full completeness.*

² $\beta(\mu^{-1})$ will generally be distinct in form from $\alpha(\lambda)$, although the second line in (7) coincides upon the substitution of λ for μ^{-1} with (4), since the integrand in (5) is $p(x_1, \dots, x_n | H_2)$ whereas that in (2), regarded as a constraint in the direct process, is $p(x_1, \dots, x_n | H_1)$.

AN INEQUALITY DUE TO H. HORNICH

BY Z. W. BIRNBAUM AND HERBERT S. ZUCKERMAN

University of Washington

H. Hornich¹ proved a theorem on the average risk of the sum of equal insurance policies. It seems of interest to note that when translated from its actuarial formulation into the terminology of the calculus of probabilities this theorem becomes an inequality for mean deviations of random variables, and to present it with a concise proof in non-actuarial language.

Let x be a random variable with a symmetrical probability distribution, $D_1 = E(|x|)$ its mean deviation, x_1, x_2, \dots, x_n independent repetitions of x , and $D_n = E(|x_1 + x_2 + \dots + x_n|)$ the mean deviation of $x_1 + x_2 + \dots + x_n$. Then D_n fulfills the inequality

$$(1) \quad D_n \geq \frac{D_1 n}{2^{n-1}} \binom{n-1}{\left[\frac{n}{2}\right]}$$

where $\left[\frac{n}{2}\right]$ denotes the greatest integer $\leq \frac{n}{2}$. If the distribution of x is not symmetrical but $E(x) = 0$, the inequality becomes

$$(2) \quad D_n \geq \frac{D_1 n}{2^n} \binom{n-1}{\left[\frac{n}{2}\right]}.$$

The proof will be given for a continuous random variable but it clearly holds quite generally. If $f(x)$ is the probability density of x , and $E(x) = 0$, then one has

$$(3) \quad D_1 = \int_{-\infty}^{+\infty} |x| f(x) dx = 2 \int_0^{\infty} x f(x) dx.$$

In the expression for D_n , the integration over the entire n -space (x_1, x_2, \dots, x_n) may be performed by integrating separately over each of the 2^n "octants" which correspond to the different combinations of signs of the coordinates, and thus one obtains, for a symmetrical distribution, the estimates

$$\begin{aligned} D_n &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \left| \sum_{i=1}^n x_i \right| \prod_{i=1}^n f(x_i) \prod_{i=1}^n dx_i \\ &= \sum_{\epsilon_1=\pm 1, \epsilon_2=\pm 1, \dots, \epsilon_n=\pm 1} \int_{\text{sgn } x_1=\epsilon_1} \int_{\text{sgn } x_2=\epsilon_2} \dots \int_{\text{sgn } x_n=\epsilon_n} \\ &= \sum_{s=0}^n \binom{n}{s} \int_{\substack{\text{sgn } x_1=\text{sgn } x_2=\dots=\text{sgn } x_s=-1 \\ \text{sgn } x_{s+1}=\text{sgn } x_{s+2}=\dots=\text{sgn } x_n=+1}} \dots \geq 2^{\left[\frac{(n-1)}{2}\right]} \sum_{s=0}^{\left[\frac{(n-1)}{2}\right]} \binom{n}{s} \int_{\substack{\text{sgn } x_1=\text{sgn } x_2=\dots=\text{sgn } x_s=-1 \\ \text{sgn } x_{s+1}=\text{sgn } x_{s+2}=\dots=\text{sgn } x_n=+1}} \dots \end{aligned}$$

¹ HANS HORNICH, "Zur theorie des Risikos," *Monatsh. Math. Phys.*, Vol. 50 (1941), pp. 142-150.

$$\begin{aligned}
&\geq 2^{\lfloor (n-1)/2 \rfloor} \binom{n}{s} \int \int \cdots \int_{\substack{\text{sgn } x_1 = \text{sgn } x_2 = \cdots = \text{sgn } x_s = -1 \\ \text{sgn } x_{s+1} = \text{sgn } x_{s+2} = \cdots = \text{sgn } x_n = +1}} \left(\sum_{i=s+1}^n x_i - \sum_{i=1}^s x_i \right) \prod_{i=1}^n f(x_i) \prod_{i=1}^n dx_i \\
&= 2^{\lfloor (n-1)/2 \rfloor} \binom{n}{s} (n-2s) \frac{D_1}{2} \cdot \frac{1}{2^{n-1}} = \frac{D_1}{2^{n-1}} \left\{ n^{\lfloor (n-1)/2 \rfloor} \binom{n}{s} - 2^{\lfloor (n-1)/2 \rfloor} s \binom{n}{s} \right\} \\
&= \frac{D_1 n}{2^{n-1}} \left\{ \sum_{s=0}^{\lfloor (n-1)/2 \rfloor} \binom{n}{s} - 2^{\lfloor (n-1)/2 \rfloor} \binom{n-1}{s} \right\} \\
&= \frac{D_1 n}{2^{n-1}} \left\{ \left(\left\lfloor \frac{n-1}{2} \right\rfloor \right) + \sum_{s=0}^{\lfloor (n-1)/2 \rfloor} \left(\binom{n}{s} - \binom{n-1}{s} \right) - \sum_{s=0}^{\lfloor (n-1)/2 \rfloor} \binom{n-1}{s} \right\} \\
&= \frac{D_1 n}{2^{n-1}} \left\{ \left(\left\lfloor \frac{n-1}{2} \right\rfloor \right) + \sum_{s=1}^{\lfloor (n-1)/2 \rfloor} \binom{n-1}{s-1} - \sum_{s=0}^{\lfloor (n-1)/2 \rfloor} \binom{n-1}{s} \right\} \\
&= \frac{D_1 n}{2^{n-1}} \left\{ \left(\left\lfloor \frac{n-1}{2} \right\rfloor \right) - \left(\left\lfloor \frac{n-1}{2} \right\rfloor - 1 \right) \right\} \\
&= \frac{D_1 n}{2^{n-1}} \left(\left\lfloor \frac{n-1}{2} \right\rfloor \right) = \frac{D_1 n}{2^{n-1}} \binom{n-1}{\lfloor \frac{n-1}{2} \rfloor}
\end{aligned}$$

If x is not symmetrical but $E(x) = 0$, we consider the random variable x' with the probability density $g(x') = f(-x')$, and the random variable $y = x + x'$. In view of (3) we find

$$\begin{aligned}
E(|y|) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x+x'| f(x) g(x') dx dx' \geq 2 \int_0^{\infty} \int_0^{\infty} (x+x') f(x) g(x') dx dx' \\
&= 2 \int_0^{\infty} g(x') \int_0^{\infty} x f(x) dx dx' + 2 \int_0^{\infty} f(x) \int_0^{\infty} x' g(x') dx' dx \\
&= E(x) \left\{ - \int_0^{\infty} g(-x) dx + \int_0^{\infty} f(x) dx \right\} = E(x).
\end{aligned}$$

Let x_1, x_2, \dots, x_n and x'_1, x'_2, \dots, x'_n be independent repetitions of x and x' , respectively, and $y_i = x_i + x'_i$. Since y has a symmetrical distribution, an application of (1) gives

$$\begin{aligned}
E(|x_1 + x_2 + \cdots + x_n|) &= \frac{1}{2} \{ E(|x_1 + \cdots + x_n|) + E(|x'_1 + \cdots + x'_n|) \} \\
&\geq \frac{1}{2} E(|x_1 + x'_1 + \cdots + x_n + x'_n|) = \frac{1}{2} E(|y_1 + \cdots + y_n|) \\
&\geq \frac{1}{2} \frac{E(|y|) n}{2^{n-1}} \binom{n-1}{\lfloor \frac{n}{2} \rfloor} \geq \frac{E(|y|) n}{2^n} \binom{n-1}{\lfloor \frac{n}{2} \rfloor}
\end{aligned}$$

An application of Stirling's formula shows that the right hand sides in (1) and (2) are of the order of magnitude of \sqrt{n} .

NOTE ON A LEMMA

BY ABRAHAM WALD

Columbia University

In a previous paper on the power function of the analysis of variance test¹, the author stated the following lemma (designated there as Lemma 2):

LEMMA 2. Let v_1, \dots, v_k be k normally and independently distributed variates with a common variance σ^2 . Denote the mean value of v_i by α_i ($i = 1, \dots, k$) and let $f(v_1, \dots, v_k, \sigma)$ be a function of the variables v_1, \dots, v_k and σ which does not involve the mean values $\alpha_1, \dots, \alpha_k$. Then, if the expected value of $f(v_1, \dots, v_k, \sigma)$ is equal to zero, $f(v_1, \dots, v_k, \sigma)$ is identically equal to zero, except perhaps on a set of measure zero.

In the paper mentioned above it was intended to state this lemma for bounded functions $f(v_1, \dots, v_k)$ and the lemma was used there only in a case where $f(v_1, \dots, v_k)$ is bounded. Through an oversight this restriction on $f(v_1, \dots, v_k)$ was not stated explicitly.² The published proof of Lemma 2 is adequate if $f(v_1, \dots, v_k)$ is assumed to be bounded. From the fact that the moments of a multivariate normal distribution determine uniquely the distribution it is concluded there that if for any set (r_1, \dots, r_k) of non-negative integers

$$(1) \quad \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} v_1^{r_1} \dots v_k^{r_k} f(v_1, \dots, v_k) e^{-\frac{1}{2}(v_1 - \alpha_1)^2} dv_1 \dots dv_k = 0$$

identically in the parameters $\alpha_1, \dots, \alpha_k$ then $f(v_1, \dots, v_k)$ must be equal to zero except perhaps on a set of measure zero. This conclusion is obvious if $f(v_1, \dots, v_k)$ is bounded. In fact, from (1) and the boundedness of $f(v_1, \dots, v_k)$ it follows that there exists a finite value A such that

$$\varphi(v_1, \dots, v_k) = \frac{1}{(2\pi)^{k/2}} \left[1 - \frac{1}{A} f(v_1, \dots, v_k) \right] e^{-\frac{1}{2}(v_1 - \alpha_1)^2}$$

is a probability density function with moments equal to those of the normal distribution

$$\psi(v_1, \dots, v_k) = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}(v_1 - \alpha_1)^2}.$$

Hence $f(v_1, \dots, v_k)$ must be equal to zero except perhaps on a set of measure zero. However, this conclusion is not so immediate if no restriction is imposed on $f(v_1, \dots, v_k)$ except that

$$(2) \quad \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} |f(v_1, \dots, v_k)| e^{-\frac{1}{2}(v_1 - \alpha_1)^2} dv_1 \dots dv_k < \infty$$

for all values of the parameters $\alpha_1, \dots, \alpha_k$. It is the purpose of this note to prove this. In other words, we shall prove the following proposition:

¹ A. WALD, "On the power function of the analysis of variance test," *Annals of Math Stat.*, Vol. 13 (1942), pp. 434.

² I wish to thank Prof. J. Neyman for calling my attention to this omission.

PROPOSITION I. If (2) holds for all values of the parameters $\alpha_1, \dots, \alpha_k$ and if for any set (r_1, \dots, r_k) of non-negative integers equation (1) holds identically in $\alpha_1, \dots, \alpha_k$, then $f(v_1, \dots, v_k)$ must be equal to zero except perhaps on a set of measure zero.

On the basis of Proposition I and the arguments given on p. 438 of the paper mentioned before, it can be seen that restriction (2) on the function $f(v_1, \dots, v_k)$ is sufficient for the validity of Lemma 2.

To prove Proposition I, we shall first show that the following lemma holds.

LEMMA A. If $h(v_1, \dots, v_k)$ is a probability density function and if

$$(3) \quad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(v_1, \dots, v_k) e^{\delta \sum |v_i|} dv_1 \cdots dv_k < \infty$$

for some $\delta > 0$, then the problem of moments is determined for the moments of the distribution $h(v_1, \dots, v_k)$.

This lemma was proved by G. H. Hardy for $k = 1$.³ I shall prove it for $k > 1$. Since

$$(4) \quad \sum_{n=0}^{\infty} \frac{\delta^{2n} (\sum |v_i|)^{2n}}{(2n)!} < e^{\delta \sum |v_i|}$$

we obtain from (3)

$$(5) \quad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(v_1, \dots, v_k) \left[\sum_{n=0}^{\infty} \frac{\delta^{2n} (\sum |v_i|)^{2n}}{(2n)!} \right] dv_1 \cdots dv_k < \infty.$$

Hence

$$(6) \quad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(v_1, \dots, v_k) \left[\sum_{n=0}^{\infty} \frac{\delta^{2n} \left(\sum_{i=1}^k v_i^{2n} \right)}{(2n)!} \right] dv_1 \cdots dv_k < \infty.$$

Denote the $2n$ th moment of v_i by $\mu_{2n}^{(i)}$. Because of (3) the moments $\mu_{2n}^{(i)}$ are finite. Furthermore, denote $\sum_{i=1}^k \mu_{2n}^{(i)}$ by λ_{2n} . Then we obtain from (6)

$$(7) \quad \sum_{n=0}^{\infty} \frac{\delta^{2n} \lambda_{2n}}{(2n)!} < \infty.$$

From (7) it follows that

$$(8) \quad \limsup_{n \rightarrow \infty} \frac{\delta^{2n} \lambda_{2n}}{(2n)!} < 1.$$

Hence

$$(9) \quad \limsup_{n \rightarrow \infty} \left(\frac{\delta^{2n} \lambda_{2n}}{(2n)!} \right)^{1/2n} \leq 1.$$

³ See for instance, SHOHAT and TAMARKIN, "The problem of moments," *Math. Surveys* No. 1, Amer. Math. Soc., New York, 1943, p. 20.

Since according to Stirling's formula

$$\lim_{n \rightarrow \infty} (2n)! / (2n)^{2n} e^{-2n} \sqrt{4\pi n} = 1$$

we obtain from (9)

$$(10) \quad \limsup_{n \rightarrow \infty} \frac{\delta \lambda_{2n}^{1/2n}}{2ne^{-1}} \leq 1.$$

Taking reciprocals we obtain

$$(11) \quad \liminf_{n \rightarrow \infty} \frac{2n \lambda_{2n}^{-1/2n}}{e\delta} \geq 1$$

or

$$(12) \quad \liminf_{n \rightarrow \infty} n \lambda_{2n}^{-1/2n} \geq \frac{e\delta}{2} > 0.$$

But (12) implies the existence of a positive value ρ so that

$$(13) \quad \lambda_{2n}^{-1/2n} \geq \frac{\rho}{n} \quad (n = 1, 2, \dots, \text{ad inf.})$$

From (13) it follows that

$$(14) \quad \sum_{n=1}^{\infty} \lambda_{2n}^{-1/2n} = \infty.$$

But (14) is Carleman's sufficient condition for the determinateness of the problem of moments. Hence Lemma A is proved.

On the basis of Lemma A we can prove Proposition I as follows: From (2) we obtain

$$(15) \quad \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} |f(v_1, \dots, v_k)| e^{-\frac{1}{2} \Sigma v_i^2 + \Sigma \alpha_i v_i} dv_1 \dots dv_k < \infty$$

for all values $\alpha_1, \dots, \alpha_k$. Let $f_1(v) = f(v)$ for all points $v = (v_1, \dots, v_k)$ for which $f(v) \geq 0$, and $f_1(v) = 0$ for all points v for which $f(v) < 0$. Similarly, let $f_2(v) = -f(v)$ for all points v for which $f(v) \leq 0$, and $f_2(v) = 0$ for all points v for which $f(v) > 0$. Then $f_1(v)$ and $f_2(v)$ are non-negative functions and

$$(16) \quad f(v) = f_1(v) - f_2(v).$$

From (15) it follows that

$$(17) \quad \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_1(v) e^{-\frac{1}{2} \Sigma v_i^2 + \Sigma \alpha_i v_i} dv_1 \dots dv_k < \infty$$

and

$$(18) \quad \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_2(v) e^{-\frac{1}{2} \Sigma v_i^2 + \Sigma \alpha_i v_i} dv_1 \dots dv_k < \infty.$$

Let

$$(19) \quad f_j^*(v) = f_j(v)e^{-\frac{1}{2}v_j^2} \quad (j = 1, 2).$$

Now we shall show that for any positive values β_1, \dots, β_k

$$(20) \quad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_j^*(v_1, \dots, v_k) e^{\beta_1|v_1| + \cdots + \beta_k|v_k|} dv_1 \cdots dv_k < \infty.$$

In fact, consider the 2^k sets (a_1, \dots, a_k) where $a_i = \pm 1$ ($i = 1, \dots, k$). Denote by $R_{a_1 \dots a_k}$ the subset of the k -dimensional Cartesian space which consists of all points $v = (v_1, \dots, v_k)$ for which v_i is either zero or signum $v_i = \text{signum } a_i$ ($i = 1, \dots, k$). Putting $\alpha_i = a_i \beta_i$, it follows from (17) and (18) that

$$(21) \quad \int_{R_{a_1 \dots a_k}} f_j^*(v_1, \dots, v_k) e^{\beta_1|v_1| + \cdots + \beta_k|v_k|} dv_1 \cdots dv_k < \infty.$$

Since (21) holds for any of the 2^k sets $R_{a_1 \dots a_k}$, equation (20) is proved.

From (1) it follows that

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} v_1^{r_1} \cdots v_k^{r_k} [f_1^*(v_1, \dots, v_k) - f_2^*(v_1, \dots, v_k)] dv_1 \cdots dv_k = 0,$$

for all non-negative integers r_1, \dots, r_k . Hence, because of (21) and Lemma A we see that

$$(22) \quad f_1^*(v_1, \dots, v_k) = f_2^*(v_1, \dots, v_k),$$

except perhaps on a set of measure zero. From (22) it follows that

$$f(v_1, \dots, v_k) = f_1(v_1, \dots, v_k) - f_2(v_1, \dots, v_k) = 0,$$

except perhaps on a set of measure zero. Hence Proposition I is proved

A NOTE ON SKEWNESS AND KURTOSIS

BY J. ERNEST WILKINS, JR.

University of Chicago

It is the purpose of §1 of this paper to prove the following inequality:

$$(1) \quad \alpha_4 \geq \alpha_3^2 + 1.$$

This inequality seems to have first been stated by Pearson¹. The inequality also follows from a result appearing in the thesis of Vatnsdal. Here we give a proof based on the theory of quadratic forms which seems to be more direct and more elementary than either of the previous proofs.

¹ "Mathematical contributions to the theory of evolution, XIX; second supplement to a memoir on skew variation," *Phil. Trans. Roy. Soc. (A)*, Vol. 216 (1916), p. 432.

The inequality (1) obviously shows that $\alpha_4 \geq 1$. It is then natural to ask for an upper bound for α_4 . In §2 we shall show that there is no universal upper bound (independent of the number N of quantities in the distribution) for α_3 . In fact we find the actual dependence of the maximum possible value of α_3 as a function of N . The form of this function seems to be known but not to have been rigorously proved before. It then follows from (1) that there is no universal upper bound for α_4 .

1. The inequality (1). Let us consider the quadratic form

$$\begin{aligned} G(a, b, c) &= \nu_0 a^2 + 2\nu_1 ab + 2\nu_2 ac + \nu_3 b^2 + 2\nu_3 bc + \nu_4 c^2 \\ &= N^{-1} \Sigma(a + xb + x^2 c)^2. \end{aligned}$$

It follows that $G(a, b, c)$ is a positive semi-definite quadratic form. In fact, if there are at least three distinct values of x , then $G(a, b, c)$ is a positive definite form. Consequently, its discriminant

$$\begin{vmatrix} \nu_0 & \nu_1 & \nu_2 \\ \nu_1 & \nu_2 & \nu_3 \\ \nu_2 & \nu_3 & \nu_4 \end{vmatrix}$$

must be non-negative. There is no loss of generality in supposing that $\nu_1 = 0$, $\nu_2 = 1$, in which case we find that

$$\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & \alpha_3 \\ 1 & \alpha_3 & \alpha_4 \end{vmatrix} \geq 0.$$

Expanding the determinant, we get the inequality (1).

We remark that equality holds in (1) if and only if there are only two distinct values of x .

2. The maximum value of α_3 . It is clear that this maximum will be N^{-1} times the maximum value of the function $f(x) = \Sigma x^3$ on the bounded closed set consisting of those points x for which $g(x) = \Sigma x^2 = N$ and $h(x) = \Sigma x = 0$. According to the Lagrange multiplier rule, this latter maximum is obtained as follows. Let $F(x) = f(x) - \lambda g(x) - \mu h(x)$. Then the maximizing point satisfies the relations

$$F_{x_i} = 3x_i^2 - 2\lambda x_i - \mu = 0, \quad \Sigma x^2 = N, \quad \Sigma x = 0.$$

The equations $\Sigma F_x = 0$, $\Sigma x F_x = 0$ shows that $\mu = 3$, $f_{\max} = 2N\lambda/3$. Solving the equation $F_{x_i} = 0$ gives

$$(2) \quad x_i = [\lambda + e_i(\lambda^2 + 9)^{1/2}]/3,$$

where $e_i = \pm 1$. For these values of x , we shall have $h(x) = 0$, $g(x) = N$ if and only if

$$\lambda = -(\lambda^2 + 9)^{1/2} N^{-1} \Sigma e_i.$$

Therefore λ has the sign opposite to that of Σe , and

$$\lambda^2[N^2 - (\Sigma e)^2] = 9(\Sigma e)^2.$$

It follows that $\Sigma e \neq \pm N$, and that

$$(3) \quad \begin{aligned} \lambda &= -3\Sigma e/[N^2 - (\Sigma e)^2]^{\frac{1}{2}}, \\ f_{\max} &= -2N\Sigma e/[N^2 - (\Sigma e)^2]^{\frac{1}{2}}. \end{aligned}$$

We have still not obtained the maximum, however, since the minimum will also satisfy all of the relations deduced above. We distinguish the maximum from the other critical values by examining the function

$$\theta(\Sigma) = -2N\Sigma/(N^2 - \Sigma^2)^{\frac{1}{2}}.$$

Since $\Sigma e \neq \pm N$, $e_i = \pm 1$, it is clear that $N - 2 \geq \Sigma e \geq 2 - N$. We therefore consider $\theta(\Sigma)$ on the interval $(2 - N, N - 2)$. We find that

$$d\theta/d\Sigma = -2N^3/(N^2 - \Sigma^2)^{3/2} < 0,$$

so that θ is a decreasing function of Σ on the interval indicated. Its maximum value will therefore occur when $\Sigma = 2 - N$, and this maximum value will be

$$\theta(2 - N) = N(N - 2)/(N - 1)^{\frac{1}{2}}.$$

The value $\Sigma e = 2 - N$ occurs only when one of the e_i , say e_1 , is equal to $+1$ and all the rest are equal to -1 . Then we find from (3) and (2) that

$$(4) \quad \begin{aligned} \lambda &= 3(N - 2)/2(N - 1)^{\frac{1}{2}}, \\ x_1 &= (N - 1)^{\frac{1}{2}}, \quad x_2 = x_3 = \cdots = x_N = -(N - 1)^{-\frac{1}{2}}, \\ \alpha_3 &= f(x)/N = (N - 2)/(N - 1)^{\frac{1}{2}}. \end{aligned}$$

Since the maximum value of α_3 given by (4) approaches ∞ with N , it follows that there is no universal upper bound for α_3 . More precisely, the quantity α_3 can be made as large as desired by choosing N large enough and then picking x_i as in the last paragraph. Since there is no universal upper bound for α_3 , it is clear from (1) that there is no universal upper bound for α_4 . It would probably be possible, although rather difficult, to derive an explicit bound for α_4 as a function of N by using the methods employed above for α_3 .

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of general interest

Personal Items

Mr. Carl A. Bennett is doing war research on a project at the University of Chicago.

Dr. George W. Brown, formerly Research Associate at Princeton University, is now at the RCA Laboratories in Princeton, New Jersey.

Professor Harry Carver is on leave of absence from the University of Michigan to do Operations Analysis work with the Army Air Forces.

Mr. George B. Dantzig is now Principal Statistician of the Army Air Forces Statistical Control Division at the Pentagon Building in North Arlington, Virginia.

Assistant Professor Preston C. Hammer of Oregon State College has taken a leave of absence to aid in setting up statistical methods of quality control for the Lockheed Aircraft Corporation of Burbank, California.

Dr. Tjalling Koopmans is now associated with the Cowles Commission for Research in Economics at the University of Chicago.

Dr. Jerome C. R. Li is now an instructor in mathematics at Queens College in Flushing, New York.

Associate Professor Joe Livers of Montana State College has been granted a leave of absence for the summer and fall terms to study at the University of Michigan.

Assistant Professor Eugene Lukacs of Illinois College, Jacksonville, Illinois, has accepted an associate professorship in mathematics at Berea College, Berea, Kentucky.

Mr. R. I. Piper, who has been on leave from the Southern California Telephone Company to do war research at the California Institute of Technology, has returned to the Telephone Company as Plant Staff Assistant.

Dr. Henry Scheffé, formerly lecturer in mathematics at Princeton University, has been appointed to an assistant professorship in mathematics at Syracuse University.

Assistant Professor H. M. Schwartz of the University of Idaho has been appointed research fellow at the Bartol Research Foundation in Swarthmore, Pennsylvania.

New Members

The following persons have been elected to membership in the Institute:

- Bower, O. K.** Ph.D. (Illinois) Asso. in Dept. of Math., Univ. of Illinois. 505 W. John, Champaign, Ill.
- Breden, Robert E.** B.S. (Kansas State College) Personnel Technician, Statistical Services, Technical Section of the War Dept., 270 Madison Ave., New York, N. Y.
- Brixey, John C.** Ph.D. (Chicago) Asso. Prof. of Math., Univ. of Oklahoma. 927 S. Pickard, Norman, Okla.
- Brown, Arthur B.** Ph.D. (Harvard) Asst. Prof. of Math., Queens College, Flushing, N. Y.
- Bruyere, Martha** (Mrs. Paul T.) Stat., U. S. Public Health Service, Div. of Venereal Diseases, Gaithersburg, Md.

- Bruyere, Paul T.** M P.H. (Yale) Stat., U. S. Public Health Service, Bldg. T6, Bethesda, Md.
- Carter, Gerald C.** Ph D. (Purdue) Dept Head, Naval Training School, Purdue Univ. 530 Garfield St., W. Lafayette, Ind.
- Casanova, Teobaldo.** Ph D. (New York) Res. Stat., Univ. of Puerto Rico, Rio Piedras, P. R.
- Chances, Ralph.** B B.S. (C C.N.Y.) Stat., Industrial Surveys Co., 347 Madison Ave., New York, N. Y.
- Cody, Donald D.** A B (Harvard) Res. Math., Res. Lab., Indianapolis Naval Ordnance Plant, Indianapolis, Ind.
- Duvall, George E.** Asst. Physicist, UCDWR, U. S. Navy Radio and Sound Lab., San Diego 52, Calif
- Ellis, Wade.** Ph.D (Michigan) Special Instr. in Math., Univ. of Michigan. 921 Woodlawn, Ann Arbor, Mich.
- Field, Robert W.** Ph.D (Illinois) Asso. Prof. of Industrial Engineering, Purdue Univ., Lafayette, Ind.
- File, Quentin W.** Ph D. (Purdue) Instr. of Elec. Wiring, Purdue Naval Training School. Physics Bldg., Purdue Univ., Lafayette, Ind
- Freeman, Albert M.** Dir. Math. Lab., Boston Fiduciary & Research Associates. Neck Rd., Tiverton, R. I
- Gerlough, Daniel L.** B.S (Calif Inst. of Tech.) Quality Control Engineer, Plomb Tool Co., Box 3519, Terminal Annex, Los Angeles 54, Calif
- Germond, Hallett H.** Ph D. (Wisconsin) Asso. Prof. of Math., Univ. of Florida, Gainesville, Fla (On leave).
- Ghormley, Glen E.** Stat. Analyst, Lockheed Aircraft Corp. 189 N. Chester Ave., Pasadena 4, Calif.
- Grant, Eugene L.** A M. (Columbia) Prof. of Economics of Engineering, Stanford Univ., Calif.
- Gunlogson, L. S.** B B.A. (Minnesota) Ensign, USNR 3616 18th Ave., So., Minneapolis 7, Minn
- Hart, Alex L.** Ph D. (Minnesota) Asst. to the Dir., Res. Dept., Eastern Air Lines, Inc. 141-24 79 Ave., Flushing, L. I., N. Y.
- Kefferstan, William F.** Mgr., Economic Research, Boston Fiduciary & Research Associates, 50 Congress St., Boston, Mass
- Lyons, Will.** B Sc (Bucknell) Economist, Gen. Statistics Staff, W.P.B 2027 Park Rd., N W, Washington 10, D. C.
- McBee, Ethelyne L.** M A (Columbia) Stat., Dept. of Agric 2126 N. Stafford St., Arlington, Va
- McIntyre, Donald P.** M A. (Toronto) Meteorologist in Charge, Prince George Airport Box 296, Prince George, B. C., Canada.
- Moss, Judith.** B A (Vassar) Jr. Math., Stat. Res. Group, Div. of War Research, Columbia Univ. 319 St. John's Pl., Brooklyn 17, N. Y.
- Oosterhof, Willis M.** M.A (Michigan) Stat., Mich. State Dept. of Social Welfare 811 Hackett St., Ionia, Mich
- Piper, Robert I.** B A (Montana) Res. Asso. Room 204, Astrophysics Bldg., Calif. Inst. of Tech., Pasadena 4, Calif.
- Robbins, Herbert E.** Ph D (Harvard) Lt., USNR. Post Graduate School, Annapolis, Md.
- Seeley, Sherwood B.** Ch E. (New York) Dir., Res. and Tech. Div., Joseph Dixon Crucible Co., 167 Wayne St., Jersey City 3, N. J.
- Simpson, Tracy W.** E.E (Armour Inst. of Tech.) Sales Promotion Mgr., Marchant Calculating Machine Co. 2903 Forest Ave., Berkeley, Calif
- Smith, Edward S.** Ph D (Virginia) Prof. of Math., Univ. of Cincinnati, Cincinnati 21, Ohio
- Waksberg, Joseph.** B S. (C C.N.Y.) Economic Analyst, Bureau of the Census. 1422 Saratoga Ave., N E, Washington 18, D. C.

REPORT ON THE WELLESLEY MEETING OF THE INSTITUTE

The Seventh Summer Meeting of the Institute of Mathematical Statistics was held at Wellesley College, Wellesley, Mass., on Saturday and Sunday, August 12 and 13, 1944, in conjunction with the meetings of the Mathematical Association of America and the American Mathematical Society. The following 51 members of the Institute attended the meeting:

T. W. Anderson, H. E. Arnold, K. J. Arnold, L. A. Aroian, A. L. Bailey, J. L. Barnes, C. I. Bliss, A. H. Bowker, B. H. Camp, C. W. Churchman, W. G. Cochran, T. E. Cope, J. H. Curtiss, W. E. Deming, P. S. Dwyer, Will Feller, C. D. Ferris, R. M. Foster, H. A. Freeman, Henry Goldberg, E. J. Gumbel, P. R. Halmos, Harold Hotelling, Truman Kelley, L. R. Klein, Myra Levine, John Mandel, J. W. Mauchly, Richard. v. Mises, E. B. Mode, Vaclav Myslivec, P. M. Neurath, M. L. Norden, C. O. Oakley, P. S. Olmstead, Edward Paulson, Wm. Reitz, S. L. Robinson, F. E. Satterthwaite, Henry Scheffé, W. A. Shewhart, Andrew Sobszyk, H. W. Steinhaus, Marian Torrey, Mary Torrey, A. W. Tucker, J. W. Tukey, Abraham Wald, R. M. Walter, Elizabeth Wilson, Jacob Wolfowitz.

The first session was held jointly with the Mathematical Association and consisted of a Symposium on "Potential Opportunities for Statisticians and the Teaching of Statistics." The President of the Institute, Dr. W. A. Shewhart, presided. The principal addresses were made by Dr. Shewhart and Professor Harold Hotelling. Remarks were also made, upon invitation of the Chairman, by Prof. Milton de Silva Rodrigues of Sao Paulo University in Brazil who is spending a year in studying methods of teaching statistics in this country, and by Dr. Vaclav Myslivec, Czechoslovak Delegate to the United Nations Interim Commission on Food and Agriculture. A lively discussion was under way when time forced the conclusion of the meeting. There was continued discussion by a smaller group for some time afterwards.

Professor B. H. Camp acted as Chairman at the Sunday morning session, a contributed papers session held jointly with the Association. The following papers were presented:

1. *Statistical Tests Based on Permutations of the Observations.*
A. Wald and J. Wolfowitz, Columbia University.
2. *Error Control in Matrix Calculation.*
F. E. Satterthwaite, Aetna Life Insurance Co.
3. *On Cumulative Sums of Random Variables.*
A. Wald, Columbia University.
4. *The Approximate Distribution of the Mean and of the Variance of Independent Variables.*
P. L. Hsu, National University of Peking. (Introduced, and presented, by W. Feller, Brown University).
5. *Ranges and Midranges.*
E. J. Gumbel, New School for Social Research.
6. *Statistics of Sensitivity Data, II. Preliminary report.*
C. W. Churchman, Frankford Arsenal and Benjamin Epstein, Westinghouse Electric and Manufacturing Co.

President Shewhart presided at the Sunday afternoon session. The following invited addresses were given:

1. *The Problem on Tolerance Limits.*

Lt. J. H. Curtiss, USNR.

2. *Some Improvements in Weighing and Other Experimental Techniques.*

H. Hotelling, Columbia University.

A business meeting was held at the conclusion of the Sunday afternoon session. The Secretary-Treasurer made a brief report dealing with (1) the financial condition of the Institute and (2) the membership growth of the Institute. The President, reporting for the Editor, indicated a need for more papers for the next two or three issues of the Annals. The Institute, after some discussion, then passed two Amendments to the Constitution and four Amendments to the By-Laws. These Amendments are listed in the following section. A resolution thanking the officials of Wellesley College was passed.

A dinner for the three mathematical organizations was held Sunday evening. Addresses were made by Captain Mildred H. McAfee and Professor Marshall H. Stone. Later in the evening there was a musicale featuring David Barnett.

P. S. DWYER

Secretary

AMENDMENTS TO THE CONSTITUTION AND BY-LAWS OF THE INSTITUTE

The following Amendments to the Constitution and By-laws were passed at the business meeting at Wellesley College on August 13, 1944. The votes of all voting members who sent ballots to the Secretary-Treasurer prior to the time of the meeting were counted in the balloting. The Amendments as adopted are identical with the proposed Amendments which were placed in the hands of the membership in July:

Amendments to Constitution

1. Article III. 3. The first sentence, which was

"The Institute shall have a Committee on Membership composed of three Fellows." shall be revised to read:

"The Institute shall have a Committee on Membership composed of a Chairman and three Fellows."

2. Article IV. 3. The first two sentences, which were:

"The Committee on Membership shall hold a meeting immediately after the annual meeting of the Institute. Further meetings of the Committee may be held from time to time at the call of the Chairman or any member of the Committee provided notice of such call and the purpose of the meeting is given to the members of the Committee by the Secretary-Treasurer at least five days before the date set therefor."

shall be revised to read:

"Meetings of the Committee on Membership may be held from time to time at the call of the Chairman or any member of the Committee provided notice of such call and the purpose of the meeting is given to the members of the Committee by the Secretary-Treasurer at least five days before the date set therefor. Committee business may also be transacted by correspondence if that seems preferable."

Amendments to By-Laws

1. Article I. 4. Add the following sentence:

"The power of election to the different grades of Membership, except the grades of Member and Junior Member, shall reside in the Board."

2. Article I. 5. which was:

"The Committee on Membership shall prepare and make available through the Secretary-Treasurer an announcement indicating the qualifications requisite for the different grades of membership."

shall have added the following sentences:

"The Committee shall review these qualifications periodically and shall make such changes in these qualifications and make such recommendations with reference to the number of grades of membership as it deems advisable. The power to elect worthy applicants to the grades of Member and Junior Member shall reside in the Committee, which may delegate this power to the Secretary-Treasurer, subject to such reservations as the Committee considers appropriate. The Committee shall make recommendations to the Board of Directors with reference to placing members in other grades of membership. The Committee shall give its attention to the question of increasing the number of applicants for membership and shall advise the Secretary-Treasurer on plans for that purpose."

3. After Article II. 1(a) Exception. Add:

"(b) Exception. Any Member or Fellow may make a single payment which will be accepted by the Institute in place of all succeeding yearly dues and which will not otherwise alter his status as a Member or Fellow. The amount of this payment will depend upon the age of this Member or Fellow and will be based upon a suitable mortality table and rate of interest, to be specified by the Board of Directors."

and

4. "(c) Exception. Any Member or Junior Member of the Institute serving, except as a commissioned officer, in the Armed Forces of the United States or of one of its allies, may upon notification to the Secretary-Treasurer be excused from the payment of dues until the January first following his discharge from the Service. He shall have all privileges of membership except that he shall not receive the Official Journal. However during the first year of his resumed regular membership he may have the right to purchase, at \$2.50 per volume, one copy of each volume of the Official Journal published during the period of his service membership."
-

ABSTRACTS OF PAPERS

Presented on August 13, 1944, at the Wellesley meeting of the Institute

1. **Statistical Tests Based on Permutations of the Observations.** A. WALD and J. WOLFOWITZ, Columbia University.

It was pointed out by Fisher that statistical tests of exact size, based on permutations of the observations, can be carried out without assuming anything about the underlying distributions except their continuity. Scheffé has proved that, for an important class of hypotheses, these tests are the only ones with regions of exact size. Tests based on permutations of the observations have been constructed by Fisher, Pitman, Welch, and the present authors. In the present paper, the authors prove a theorem on the limiting normality of the distribution, in the universe of permutations, of a class of linear forms. Application of this theorem gives the limiting normality (in the universe of permutations, of course) of the correlation coefficient, and of a statistic introduced by Pitman to test the difference between two means. The limiting distribution of the analysis of variance statistic in the universe of permutations is also obtained.

2. **Error Control in Matrix Calculation.** FRANKLIN E. SATTERTHWAIT, Aetna Life Insurance Co

The arithmetic evaluation of matrix expressions is often rather complicated. One of the causes of this is the fact that relatively minor errors (such as rounding errors) introduced in an early step may be magnified to such an extent in succeeding steps that the final result is useless. Iterative methods to meet this difficulty have been reviewed very completely by Hotelling. In this paper a different approach is taken. Conditions on the norm of a matrix are determined so that a Doolittle process will not magnify errors to more than two or three decimal places. It is then pointed out that if an approximation to the inverse of the matrix is available, most problems can be rearranged so that the required norm conditions are met. A Doolittle process may then be used to any number of decimal places with assurance that errors will not accumulate to more than a limited number of decimal places.

3. **On Cumulative Sums of Random Variables.** A. WALD, Columbia University.

Let $\{z_i\}$ ($i = 1, 2, \dots$ ad inf.) be a sequence of independent random variables each having the same distribution. Denote by Z_j the sum of the first j elements of the sequence. Let $a > 0$ and $b < 0$ be two constants and denote by n the smallest integer for which either $Z_n \geq a$ or $Z_n \leq b$. Neglecting the quantity by which Z_n may differ from a or b (this can be done if the mean value of $|z_i|$ is small), the probability that $Z_n \geq c$ for $c = a$ and $c = b$ is derived, and the characteristic function of n is obtained. The probability distribution of n when z_i is normally distributed is derived. These results have application to various statistical problems and to problems in molecular physics dealing with the random walk of particles in the presence of absorbing barriers.

4. **The Approximate Distribution of the Mean and of the Variance of Independent Variates.** P. L. HSU, National University of Peking.

Let X_k be mutually independent random variables with the same cumulative distribution function; let $E(X_k) = 0$, $E(X_k^2) = 1$ and $E(X_k^3) = \delta$. Finally put $S = n^{-1} \sum_{k=1}^n X_k$ and $\eta = n^{-1} \sum_{k=1}^n (X_k - S)^2$. The author first gives a new derivation of H. Cramer's well-known asymptotic expansions for $Pr(n^{1/2}S \leq z)$. The proof is much more elementary and

avoids in particular the use of M. Riesz' singular integrals. Instead a considerably simpler Cesaro-type kernel is used, which has first been introduced by A. C. Berry (*Trans. Amer. Math. Soc.* 49 (1941), pp. 122-136). The same method is then used to derive similar asymptotic expansions for $Pr(n^{1/2}(\eta - 1) \leq (\delta - 1)^{1/2}x)$. The method can be extended to the case of unequal components and also for the study of other functions encountered in mathematical statistics.

5. Ranges and Midranges. E. J. GUMBEL, New School for Social Research

The m th range w_m and the m th midrange v_m are defined as the difference and as the sum of the m th extreme value taken in descending magnitude ("from above") and the m th extreme value taken in ascending magnitude ("from below"). The semi-invariant generating functions $L_m(t)$ and ${}_mL(t)$ of the m th extreme values from above and below are simple generalizations of the semi-invariant generating functions of the largest and of the smallest value which have been given by R. A. Fisher and L. H. C. Tippett. If the sample size is large enough the two m th extreme values may be considered as independent variates. Then, the semi-invariant generating functions $L_w(t, m)$ and $L_v(t, m)$ of the m th range and of the m th midrange are

$$L_w(t, m) = L_m(t) + {}_mL(-t); L_v(t, m) = L_m(t) + {}_mL(t).$$

If the initial distribution is symmetrical the semi-invariant generating function of the m th range is twice the semi-invariant generating function of the m th extreme value from above. The distribution of the m th range is skew, whereas the distribution of the m th midrange is of the generalized, symmetrical, logistic type. The even semi-invariants of the m th midrange are equal to the even semi-invariants of the m th range. For increasing indices m the distributions of the m th extremes, of the m th ranges and of the m th midranges converge toward normality.

6. Statistics of Sensitivity Data, II. Preliminary report. C. W. CHURCHMAN, Frankford Arsenal, and BENJAMIN EPSTEIN, Westinghouse Electric and Manufacturing Co.

In this paper a study is made of the distribution of the first two moments of sensitivity data as functions of the sample size. The chief results are briefly these:

- (a) The distributions of \bar{x} and $\sigma_{\bar{x}}^2$ (for definition of these functions, see "On the Statistics of Sensitivity Data," by the authors in the *Annals of Mathematical Statistics*, Vol. XV, No. 1) approach normality rapidly as functions of the sample size;
- (b) \bar{x} and $\sigma_{\bar{x}}^2$ are "almost" independent even for small sample sizes, thus justifying the use of Student's ratio in tests of significance for differences between two sample means.

SOME EXTENSIONS OF THE WISHART DISTRIBUTION¹

T. W. ANDERSON AND M. A. GIRSHICK

Princeton University and Columbia University

1. Introduction. The well-known Wishart distribution is the distribution of the variances and covariances of a sample drawn from a multivariate normal population assuming that the expected value of each variate remains the same from observation to observation. For problems such as testing collinearity [1], comparing scales of measurement [2], and multiple regression in times series analysis [3], it is desirable to have the distribution of sample variances and covariances for observations, the expected values of which are not all identical. Such a distribution could be considered as a generalization to several variates of the χ'^2 (non-central χ^2) distribution, as well as a generalization of the Wishart distribution to the non-central case. In this paper we shall discuss the general problem of finding the distribution in question and shall derive this distribution for two particular cases. We shall start out with the problem in its most general form and as a result of linear transformations express the distribution as a certain multiple integral.

We can think of the expected values of the observations as defining points in a space of dimension equal to the number of variates. If these points lie on a line, the non-central Wishart distribution is essentially the Wishart distribution multiplied by a Bessel function; if the points lie in a plane, it is a Wishart distribution multiplied by an infinite series of Bessel functions. For higher dimensionality the integration of the multiple integral becomes extremely troublesome; it has not been possible yet to express the general integration in a concise form. These results are summarized precisely at the end of the paper.

2. Reduction to canonical form. Consider a set of N multivariate normal populations each of p variates. Let the i th ($i = 1, 2, \dots, p$) variate of the α th ($\alpha = 1, 2, \dots, N$) population be $x_{i\alpha}$; let the mean of this variate be

$$(1) \quad E(x_{i\alpha}) = \mu_{i\alpha} \quad (i = 1, 2, \dots, p; \alpha = 1, 2, \dots, N);$$

and let the variance-covariance matrix (of rank p) common to all N distributions be

$$|| E[(x_{i\alpha} - \mu_{i\alpha})(x_{j\alpha} - \mu_{j\alpha})] || = || \sigma_{ij} || \quad (\alpha = 1, 2, \dots, N).$$

Now consider a sample of observations $\{x_{i\alpha}\}$ one from each population.

The purpose of this paper is to find the joint distribution of the quantities

$$(2) \quad a_{ij} = \sum_{\alpha=1}^N (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j),$$

¹ The results given below were arrived at independently by the two authors. Some preliminary results were given before the Institute of Mathematical Statistics at Washington, D. C., May 6, 1944, by Girshick.

where

$$\bar{x}_i = \frac{1}{N} \sum_{\alpha=1}^N x_{i\alpha}.$$

To simplify the notation in the subsequent work we treat the quantities a_i , instead of the sample variances and covariances which are simply multiples (by $1/(N-1)$) of the a_i .

The a_i may be considered as sums of squares and cross products, for there exists a linear transformation²,

$$x'_{i\alpha} = \sum_{\beta=1}^n \theta_{\alpha\beta} x_{i\beta} \quad (i = 1, 2, \dots, p),$$

where the matrix $|| \theta_{\alpha\beta} ||$ is orthogonal (and $\theta_{\alpha 1} = \theta_{\alpha 2} = \dots = \theta_{\alpha N} = 1/\sqrt{N}$), such that

$$a_{ij} = \sum_{\alpha=1}^n x'_{i\alpha} x'_{j\alpha},$$

where $n = N - 1$ and

$$N\bar{x}_i\bar{x}_j = x'_{iN}x'_{jN}.$$

For a given α the $x'_{i\alpha}$ have a multivariate normal distribution with the same variances and covariances as the x 's and with expected values

$$E(x'_{i\alpha}) = \sum_{\beta=1}^n \theta_{\alpha\beta} \mu_{i\beta} = \mu'_{i\alpha}, \quad \text{say.}$$

Let

$$\tau_{ij} = \sum_{\alpha=1}^n \mu'_{i\alpha} \mu'_{j\alpha}.$$

Then it is clear that the τ_{ij} are the same functions of the μ 's that the a_{ij} are of the x 's, namely,

$$(3) \quad \tau_{ij} = \sum_{\alpha=1}^N (\mu_{i\alpha} - \bar{\mu}_i)(\mu_{j\alpha} - \bar{\mu}_j),$$

where

$$\bar{\mu}_i = \frac{1}{N} \sum_{\alpha=1}^N \mu_{i\alpha}.$$

Now consider the two p by p matrices

$$\Sigma = || \sigma_{ij} ||$$

and

$$T = || \tau_{ij} ||.$$

² See, for example, [4].

Let $\kappa_1^2, \kappa_2^2, \dots, \kappa_p^2$ be the real, non-negative roots of the determinantal equation

$$(4) \quad |T - \lambda \Sigma| = 0.$$

There exists a non-singular p by p matrix

$$(5) \quad \Psi = ||\psi_{ij}||$$

such that³

$$(6) \quad \Psi \Sigma \Psi' = I$$

and

$$(7) \quad \Psi T \Psi' = \begin{vmatrix} \kappa_1^2 & 0 & \cdots & 0 \\ 0 & \kappa_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \kappa_p^2 \end{vmatrix},$$

where I is the identity matrix and Ψ' is the transpose of Ψ . Suppose the rank of T is t ; then t of the roots are non-zero and $p - t$ are zero. For the sake of convenience we shall choose $\kappa_1^2, \kappa_2^2, \dots, \kappa_t^2$ to be the non-zero roots. If T is of rank t , then the means $\mu_{i\alpha}$ lie in a t dimensional sub-space of the original p dimensional space. Let us make the transformation

$$(8) \quad z_{i\alpha} = \sum_{j=1}^p \psi_{ij} x'_{j\alpha}.$$

The $z_{i\alpha}$ are normally and, because of relationship (6), independently distributed with unit variances. The mean value of $z_{i\alpha}$ is

$$E(z_{i\alpha}) = \sum_{j=1}^p \psi_{ij} \mu'_{j\alpha} = \nu_{i\alpha},$$

say. As a result of (7)

$$(9) \quad \sum_{\alpha=1}^n \nu_{i\alpha}^2 = \kappa_i^2 \quad (i = 1, 2, \dots, p),$$

$$(10) \quad \sum_{\alpha=1}^n \nu_{i\alpha} \nu_{j\alpha} = 0 \quad (i \neq j).$$

Let the new sum of squares of cross-products be

$$(11) \quad b_{ij} = \sum_{\alpha=1}^n z_{i\alpha} z_{j\alpha}.$$

We shall first find the joint distribution of the b_{ij} and then obtain the distribution of the a_{ij} by using the fact that the b_{ij} can be considered simply as a linear

³ See, for example, [5].

transformation of the a_i , in a $\frac{1}{2}p(p+1)$ dimensional space. For we can write b_{ij} as

$$(12) \quad b_{ij} = \sum_{\alpha=1}^n \sum_{h,k=1}^p \psi_{ih} \psi_{jk} x'_{h\alpha} x'_{k\alpha} = \sum_{h,k=1}^p \psi_{ih} \psi_{jk} a_{hk}.$$

The transformation (8) is performed on all variates of each observation. The next transformation, which is the one that results in the canonical form of the problem, is performed on all observations of each variate. We wish to construct the n by n matrix of this transformation

$$\Phi = || \phi_{\alpha\beta} ||$$

in the following manner: Let

$$\phi_{\eta\alpha} = \frac{\nu_{\eta\alpha}}{\kappa_{\eta}} \quad (\alpha = 1, 2, \dots, n; \eta = 1, 2, \dots, t).$$

In view of (9) and (10)

$$\sum_{\alpha=1}^n \phi_{\xi\alpha} \phi_{\eta\alpha} = \delta_{\xi\eta}, \quad (\xi, \eta = 1, 2, \dots, t)$$

where $\delta_{\xi\eta}$ is the Kronecker delta. The remaining elements in Φ are chosen in any way to make Φ orthogonal.

Now make the transformation

$$y_{i\alpha} = \sum_{\beta=1}^n \phi_{\alpha\beta} z_{i\beta} \quad (i = 1, 2, \dots, p; \alpha = 1, 2, \dots, n).$$

Because Φ is orthogonal,

$$(13) \quad b_{ij} = \sum_{\alpha=1}^n y_{i\alpha} y_{j\alpha},$$

and the y 's are independently normally distributed. By virtue of the construction of Φ and the properties of $|| \nu_{i\alpha} ||$ the expected value of each $y_{i\alpha}$ is zero except for t of the variates, namely,

$$E(y_{\eta\eta}) = \kappa_{\eta} \quad (\eta = 1, 2, \dots, t).$$

Now the problem can be put in this form: Find the distribution of b_{ij} (given by (13)) when the distribution of the y 's (in the canonical form) is

$$\frac{1}{(2\pi)^{pn/2}} e^{-\frac{1}{2} \sum_{i=1}^p \sum_{\alpha=1}^n (\nu_{i\alpha} - \kappa_i \delta_{i\alpha})^2},$$

where $\kappa_1, \kappa_2, \dots, \kappa_t$ are different from zero.

The solution of our problem can be expressed as a certain multiple integral of tp variables. Let

$$b'_{ij} = \sum_{\alpha=i+1}^n y_{i\alpha} y_{j\alpha}.$$

Since the $b'_{i,j}$ have the Wishart distribution with $n - t$ degrees of freedom (we assume $n \geq t + p$) we can write the joint distribution of the $b'_{i,j}$ and $y_{i\eta}$ ($i = 1, 2, \dots, p; \eta = 1, 2, \dots, t$) as

$$\frac{1}{2^{tp(n-t)} \pi^{\frac{1}{2}p(p-1)} \prod_{i=1}^p \Gamma(\frac{1}{2}[n-t+1-i])} |b'_{i,j}|^{\frac{1}{2}(n-p-t-1)} e^{-\frac{1}{2} \sum_{i,j=1}^p b'_{i,j}} \\ \times \frac{1}{(2\pi)^{pt/2}} e^{-\frac{1}{2} \sum_{i=1}^p \sum_{\eta=1}^t (\nu_{i\eta} - \kappa_{\eta} b_{i\eta})^2}$$

Considering the equations

$$b_{i,j} = b'_{i,j} + \sum_{\eta=1}^t y_{i\eta} y_{j\eta}$$

as a transformation of the $b'_{i,j}$, we immediately obtain the joint distribution of the b 's and the $y_{i\eta}$ ($i = 1, 2, \dots, p; \eta = 1, 2, \dots, t$) as

$$(14) \quad \frac{e^{-\frac{1}{2} \sum_{i=1}^p \kappa_i^2}}{2^{tpn} \pi^{\frac{1}{2}p(p-1)+\frac{1}{2}pt} \prod_{i=1}^p \Gamma(\frac{1}{2}[n-t+1-i])} |b_{i,j} - \sum_{\eta=1}^t y_{i\eta} y_{j\eta}|^{\frac{1}{2}(n-p-t-1)} \\ \times e^{-\frac{1}{2} \sum_{i=1}^p b_{i,i} + \sum_{\eta=1}^t \kappa_{\eta} y_{i\eta}}$$

To find the distribution of the $b_{i,j}$, we must integrate out the $y_{i\eta}$, where the range of integration is such that the matrix

$$||b_{i,j} - \sum_{\eta=1}^t y_{i\eta} y_{j\eta}||$$

is positive. For $t = 1$ or 2 we can integrate (14) and express the results in a convenient form. However, for higher values of t the integration affords considerable difficulty and has not been done for the general case. In terms of geometry the case $t = 1$ is the case in which the expected values of the observations lie on a line in the p dimensional space. In the case of $t = 2$, similarly the expected values lie in a plane in this space. Hence, we shall call these two cases the linear and planar cases, respectively.

3. The linear case. In the linear case there is one root of the equation (4) which is not equal to zero, that is, there is simply one κ in the distribution (14) and one set of y 's, namely y_{i1} ($i = 1, 2, \dots, p$). The problem is to integrate the y_{i1} over the range for which the matrix

$$||b_{i,j} - y_{i1} y_{j1}||$$

is positive; the integrand we are interested in is (dropping the subscript "1" from the κ_1 and y_1 and neglecting the part not involving the y 's)

$$(15) \quad |b_{ij} - y_i y_j|^{1(n-p-2)} e^{\kappa y_1} \prod_{i=1}^p dy_i.$$

The determinant in expression (15) can be expressed as

$$|b_{ij} - y_i y_j| = |b_{ij}| \left(1 - \sum_{i,j=1}^p b^{ij} y_i y_j \right),$$

where

$$||b^{ij}|| = ||b_{ij}||^{-1}.$$

The inverse exists because the probability is zero that $||b_{ij}||$ is singular. There is a linear transformation

$$y_i = \sum_{j=1}^p q_{ij} u_j,$$

such that

$$\sum_{i,j=1}^p b^{ij} y_i y_j = \sum_{j=1}^p u_j^2$$

and

$$\kappa y_1 = l u_1,$$

where l^2 is the one non-zero root of the equation:

$$(16) \quad |\lambda B^{-1} - \kappa^2 E_{11}| = 0,$$

where $B^{-1} = ||b^{ij}||$ and E_{11} is the matrix with unity in the upper left hand corner and zeros elsewhere. This fact is a result of the well-known theorem concerning diagonalization of pairs of quadratic forms.⁴ The Jacobian of this transformation is

$$|q_{ij}| = |b_{ij}|^{\frac{1}{2}},$$

and the range of integration is $\sum_{i=1}^p u_i^2 \leq 1$. The integrand is transformed into

$$|b_{ij}|^{1(n-p-1)} \left(1 - \sum_{i=1}^p u_i^2 \right)^{1(n-p-2)} e^{l u_1} \prod_{i=1}^p du_i.$$

Now let

$$u_1 = \sin w,$$

$$u_i = \cos w \, v_{i-1} \quad (i = 2, 3, \dots, p)$$

⁴ The transformation is the so-called "regression transformation." See Madow [6]

The Jacobian of this transformation is $\cos^p w$ and

$$1 - \sum_{i=1}^p u_i^2 = \cos^2 w \left(1 - \sum_{i=1}^{p-1} v_i^2 \right).$$

The integration is over the ranges of $\frac{\pi}{2} \leq w \leq \frac{\pi}{2}$ and

$$\sum_{i=1}^{p-1} v_i^2 \leq 1.$$

We integrate the following expression

$$|b_{i,}|^{\frac{1}{2}(n-p-1)} \left[\left(1 - \sum_{i=1}^{p-1} v_i^2 \right)^{\frac{1}{2}(n-p-2)} \prod_{i=1}^{p-1} dv_i \right] \{ \cos^{n-2} w e^{i \sin w} dw \}.$$

The integral of the quantity within the brackets is simply a Dirichlet integral [7] and its value is

$$\frac{\Gamma(\frac{1}{2}[n-p]) \pi^{\frac{1}{2}(p-1)}}{\Gamma(\frac{1}{2}[n-1])}.$$

The integral of the expression within the braces is a multiple of a Bessel function of purely imaginary argument [8, p 79]; that is,

$$\frac{\Gamma(\frac{1}{2}[n-1]) \sqrt{\pi}}{(l/2)^{\frac{1}{2}(n-2)}} I_{\frac{1}{2}(n-2)}(l).$$

Hence, the integral of (15) is

$$(17) \quad |b_{i,}|^{\frac{1}{2}(n-p-1)} \Gamma(\frac{1}{2}[n-p]) \pi^{\frac{1}{2}p} (l/2)^{-\frac{1}{2}(n-2)} I_{\frac{1}{2}(n-2)}(l).$$

Multiplying equation (16) by the determinant $|B|$ one can easily show that the non-zero root, l^2 , is simply $\kappa^2 b_{11}$. The distribution of the $b_{i,}$ in the linear case then is

$$\frac{e^{-\frac{1}{2}\kappa^2} e^{-\frac{1}{2}\sum_{i=1}^p b_{i,}}}{2^{\frac{1}{2}pn - \frac{1}{2}(n-2)} \pi^{\frac{1}{2}p(p-1)} \prod_{i=1}^{p-1} \Gamma(\frac{1}{2}[n-i])} |b_{i,}|^{\frac{1}{2}(n-p-1)} (\kappa^2 b_{11})^{-\frac{1}{2}(n-2)} I_{\frac{1}{2}(n-2)}(\kappa \sqrt{b_{11}}),$$

In §5 we shall give the distribution in terms of the original variables, namely, the $a_{i,}$.

4. The planar case. The case of two non-zero roots of equation (4) can be handled by continuing the process of integration of §3 another step. The essential problem is the integration of

$$(18) \quad |b_{i,}| - \sum_{\eta=1}^2 y_{i\eta} y_{j\eta} |^{\frac{1}{2}(n-p-3)} e^{\kappa_1 y_{11} + \kappa_2 y_{22}} \prod_{i=1}^p \prod_{\eta=1}^2 dy_{i\eta}$$

over the range of y 's for which the matrix

$$||b_{ij} - \sum_{\alpha=1}^2 y_{i\alpha} y_{j\alpha}||$$

is positive. The integration is done in two stages, first with respect to the y_{11} , then with respect to the y_{12} . Letting

$$\bar{b}_{ij} = b_{ij} - y_{11} y_{11},$$

$$\bar{n} = n - 1,$$

$$\bar{\kappa} = \kappa_1,$$

$$\bar{y}_i = y_{i1},$$

and omitting for the time being

$$(19) \quad e^{\kappa_2 y_{22}} \prod_{i=1}^p dy_{i2},$$

we can write the first stage of the integration of (18) as

$$\int ||\bar{b}_{ij} - \bar{y}_i \bar{y}_j||^{1(\bar{n}-p-2)} e^{\bar{\kappa} \bar{y}_1} \prod_{i=1}^{\bar{n}} d\bar{y}_i$$

over the range $||\bar{b}_{ij} - \bar{y}_i \bar{y}_j||$ positive. But, the only difference between this and the integration of (15) which has been shown to be (17), is that we are now writing all variables with " \sim " signs. Keeping this in mind, changing back again to our other variables of Section 4 and inserting again (19) we can write the first stage of the integration of (18) as

$$(20) \quad \Gamma(\frac{1}{2}(n-p-1))\pi^{1p} ||b_{ij} - y_{i2} y_{j2}||^{1(n-p-2)} e^{\kappa_2 y_{22}} \left[\kappa_1^2 (b_{11} - y_{12}^2) \right]^{-1(n-3)} \\ \cdot I_{1(n-3)}(\sqrt{\kappa_1^2 b_{11} - \kappa_1^2 y_{12}^2}) \prod_{i=1}^p dy_{i2}.$$

Now we must integrate (20) with respect to the y_{i2} over the range $||b_{ij} - y_{i2} y_{j2}||$ positive. The determinant in (20) can be written as

$$|b_{ij}| \left(1 - \sum_{i,j=1}^p b^{ij} y_{i2} y_{j2} \right).$$

There is a transformation

$$(21) \quad y_{i2} = \sum_{j=1}^p g_{ij} s_j,$$

such that

$$(22) \quad \sum_{i,j=1}^p b^{ij} y_{i2} y_{j2} = \sum_{j=1}^p s_j^2, \\ \kappa_1^2 y_{12}^2 = f^2 s_1^2, \\ \kappa_2 y_{22} = d_1 s_1 + d_2 s_2,$$

where f^2 is the one non-zero root of the equation

$$|\lambda B^{-1} - \kappa_1^2 E_{11}| = 0,$$

where B^{-1} and E_{11} are used as in equation (16). Since this equation is similar to (14), then $f^2 = \kappa_1^2 b_{11}$. The values of d_1 and d_2 will be considered later. This result is deduced from an extension of the theorem concerning the diagonalization of pairs of quadratic forms.⁵ The Jacobian is

$$|g_{ij}| = |b_{ij}|^{\frac{1}{2}},$$

and the range of integration is $\sum_{i=1}^p s_i^2 \leq 1$. The integrand (20) is now changed to

$$\Gamma(\tfrac{1}{2}[n - p - 1])\pi^{\frac{1}{2}p} |b_{ij}|^{\frac{1}{2}(n-p-1)} \left(1 - \sum_{i=1}^p s_i^2\right)^{\frac{1}{2}(n-p-2)} e^{d_1 s_1 + d_2 s_2} \\ \cdot \left[\frac{f^2(1 - s_1^2)}{4}\right]^{-\frac{1}{2}(n-3)} I_{\frac{1}{2}(n-3)}(\sqrt{f^2(1 - s_1^2)}) \prod_{i=1}^p ds_i.$$

Next the following transformation is made:

$$\begin{aligned} s_1 &= \sin w_1, \\ s_2 &= \cos w_1 \sin w_2, \\ s_i &= \cos w_1 \cos w_2 \cdots \cos w_{i-2} \sin w_{i-1} \quad (i = 3, 4, \dots, p). \end{aligned}$$

The Jacobian is $\cos^p w_1 \cos^{p-1} w_2$, and

$$1 - \sum s_i^2 = \cos^2 w_1 \cos^2 w_2 \left(1 - \sum_{i=1}^{p-2} v_i^2\right).$$

We now integrate

$$\begin{aligned} &\Gamma(\tfrac{1}{2}[n - p - 1])\pi^{\frac{1}{2}p} |b_{ij}|^{\frac{1}{2}(n-p-1)} \left[\left(1 - \sum_{i=1}^{p-2} v_i^2\right)^{\frac{1}{2}(n-p-2)} \prod_{i=1}^{p-2} dv_i\right] \\ (23) \quad &\cdot \left\{ \cos^{n-2} w_1 \cos^{n-3} w_2 e^{d_1 \sin w_1 + d_2 \cos w_1 \sin w_2} \left(\frac{f}{2} \cos w_1\right)^{-\frac{1}{2}(n-3)} \right. \\ &\quad \left. \cdot I_{\frac{1}{2}(n-3)}(f \cos w_1) dw_2 dw_3 \right\}. \end{aligned}$$

The integral of the expression within the square brackets is another Dirichlet integral; its value is

$$\frac{\Gamma(\tfrac{1}{2}[n - p])\pi^{\frac{1}{2}(p-2)}}{\Gamma(\tfrac{1}{2}[n - 2])}.$$

⁵ Again the "regression transformation" is used. See footnote for Section 3.

Similar to Section 3 the integration of the quantity within the braces with respect to w_2 is

$$\Gamma(\tfrac{1}{2}[n-2])\sqrt{\pi}\left(\frac{d_2 \cos w_1}{2}\right)^{-\frac{1}{2}(n-2)} I_{\frac{1}{2}(n-2)}(d_2 \cos w_1) \\ \cdot \cos^{n-2} w_1 e^{d_1 \sin w_1} \left(\frac{f \cos w_1}{2}\right)^{-\frac{1}{2}(n-2)} I_{\frac{1}{2}(n-2)}(f \cos w_1).$$

Since the range of integration of w_1 is $-\pi/2 \leq w_1 \leq \pi/2$ and since $\sin w_1$ is an odd function and $\cos w_1$ is an even function, the integral of the above expression can be transformed into an integration over the range $0 \leq w_1 \leq \pi/2$ by replacing $e^{d_1 \sin w_1}$ by $2 \sinh(d_1 \sin w_1)$. In view of the relationship between the Bessel functions of purely imaginary argument and $\sinh(d_1 \sin w_1)$ [8, p. 54] we can write the integral of the above expression as

$$\Gamma(\tfrac{1}{2}[n-2])\sqrt{\pi}\left(\frac{d_2}{2}\right)^{-\frac{1}{2}(n-2)} \sqrt{d_1} \sqrt{2\pi} \\ \times \int_0^{\pi/2} I_{\frac{1}{2}}(d_1 \sin w_1) I_{\frac{1}{2}(n-2)}(d_2 \cos w_1) I_{\frac{1}{2}(n-2)}(f \cos w_1) \sin^{\frac{1}{2}} w_1 \cos w_1 dw_1.$$

This integral can be expressed in another form by virtue of a formula in Watson's *Bessel Functions* [8, p. 377] as

$$(21) \quad \sqrt{\pi} 2^{\frac{1}{2}(n-2)} \int_0^{\pi} \frac{I_{\frac{1}{2}(n-2)}(\sqrt{d_1^2 + d_2^2 + f^2 - 2d_2 f \cos u})}{(d_1^2 + d_2^2 + f^2 - 2d_2 f \cos u)^{\frac{1}{2}(n-2)}} \sin^{n-3} u du.$$

Letting $d_1^2 + d_2^2 + f^2 = x$ and $d_2 f = y$ and using an expression formula for Bessel functions [8, p. 140] we can write (21) as

$$\sqrt{\pi} 2^{\frac{1}{2}(n-2)} \int_0^{\pi} \sum_{\gamma=0}^{\infty} \frac{(-y)^{\gamma} \cos^{\gamma} u}{\gamma!} x^{-\frac{1}{2}(n-2)+\gamma} I_{\frac{1}{2}(n-2)+\gamma}(\sqrt{x}) \sin^{n-3} u du.$$

Since the integral of $\cos^{\gamma} u \sin^{n-3} u$ where γ is odd is zero, the result of the integration (using the "duplication formula" for Γ functions and letting $\gamma = 2\omega$) is

$$\Gamma(\tfrac{1}{2}[n-2])\pi 2^{\frac{1}{2}(n-2)} \sum_{\omega=0}^{\infty} \frac{(y^2)^{\omega} x^{-\frac{1}{2}(n-2)+2\omega}}{2^{2\omega} \omega! \Gamma\left(n - \frac{1}{2} + \omega\right)} I_{\frac{1}{2}(n-2)+2\omega}(\sqrt{x}).$$

From the relationship (22) it is clear that the equation

$$(25) \quad |\kappa_i^2 \delta_{ij} - \lambda b^{ij}| = 0 \quad (\kappa_i^2 = 0 \text{ for } i = 3, 4, \dots, p)$$

is transformed by (21) into

$$(26) \quad \begin{vmatrix} f^2 + d_1^2 - \lambda & d_1 d_2 & 0 & \cdots & 0 \\ d_1 d_2 & d_2^2 - \lambda & 0 & \cdots & 0 \\ 0 & 0 & -\lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda \end{vmatrix} = 0,$$

Expression (25) is equivalent to

$$(27) \quad |k, k, b_{11} - \lambda \delta_{11}| = 0.$$

Hence, x and y^2 , which are the sum and product, respectively, of the non-zero roots of the two equivalent equations (26) and (27), are given by

$$x = f^2 + d_1^2 + d_2^2 = \kappa_1^2 b_{11} + \kappa_2^2 b_{22},$$

$$y^2 = f^2 d_2^2 = \kappa_1^2 \kappa_2^2 (b_{11} b_{22} - b_{12}^2).$$

In view of this result we can now write the integration of (23) as

$$\Gamma(\tfrac{1}{2}[n-p-1])\Gamma(\tfrac{1}{2}[n-p])\pi^p 2^{\frac{1}{2}(n-2)} |b_{11}|^{\frac{1}{2}(n-p-1)} \times \sum_{\omega=0}^{\infty} \frac{[\kappa_1^2 \kappa_2^2 (b_{11} b_{22} - b_{12}^2)]^{\omega}}{2^{2\omega} \omega! \Gamma(\frac{n-1}{2} + \omega)} \\ \cdot (\kappa_1^2 b_{11} + \kappa_2^2 b_{22})^{-\frac{1}{2}(n-2/2+2\omega)} I_{\frac{1}{2}(n-2)+2\omega}(\sqrt{\kappa_1^2 b_{11} + \kappa_2^2 b_{22}}).$$

Finally, by multiplying in what was left out of (18) we obtain the integral of (14) which is the solution to the problem as stated in the canonical form:

$$(28) \quad \frac{e^{-\frac{1}{2}(\kappa_1^2 + \kappa_2^2)} |b_{11}|^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2} \sum_{i=1}^p b_{ii}}}{2^{\frac{1}{2}pn - \frac{1}{2}(n-2)} \pi^{\frac{1}{2}p(p-1)} \prod_{i=1}^{p-2} \Gamma(\tfrac{1}{2}[n-1-i])} \\ \times \sum_{\omega=0}^{\infty} \frac{[\kappa_1^2 \kappa_2^2 (b_{11} b_{22} - b_{12}^2)]^{\omega}}{2^{2\omega} \omega! \Gamma(\frac{n-1}{2} + \omega)} (\kappa_1^2 b_{11} + \kappa_2^2 b_{22})^{-\frac{1}{2}(\frac{1}{2}(n-2)+2\omega)} \\ \times I_{\frac{1}{2}(n-2)+2\omega}(\sqrt{\kappa_1^2 b_{11} + \kappa_2^2 b_{22}}).$$

5. Final form. To answer the problem as stated originally it is necessary to make the transformation (12) and obtain the distribution in terms of the a_i , for the linear and planar cases.

It is clear that equation (25) is equivalent to

$$(29) \quad |T - \lambda A^{-1}| = 0,$$

where $A^{-1} = ||a_i||^{-1}$, for T is the transform (by (5)) of $||\kappa_i^2 \delta_i||$ and A^{-1} is the transform of $||b_i||$. The sum and product of the non-zero roots, which are the arguments of the infinite series in (24), remain unchanged

Since the quantity $\sum_{i=1}^p \kappa_i^2$ is the sum of the roots of (4) it can be expressed as

$$\sum_{i=1}^p \kappa_i^2 = \sum_{i,j=1}^p \sum_{\alpha=1}^N \sigma^{ij} (\mu_{i\alpha} - \bar{\mu}_i) (\mu_{j\alpha} - \bar{\mu}_j),$$

where $||\sigma^{ij}|| = ||\sigma_i||^{-1}$. Furthermore we have

$$|b_{11}| = |\Sigma| \cdot |a_{11}|$$

and

$$\sum_{i=1}^p b_{ii} = \text{tr} || b_{ij} || = \text{tr} (\Psi || a_{ij} || \Psi') = \sum_{i,j=1}^p \sigma^{ij} a_{ij}.$$

Moreover, the Jacobian of the transformation (12) is⁶

$$(30) \quad |J| = |\Psi|^{p+1} = |\Sigma|^{-1(p+1)}.$$

Hence, we have the following results:

Given N multivariate normal populations each of p variates with identical variance-covariance matrices $||\sigma_{ij}||$ and with expected values of the pN variates x_{ia} given by (1). Let a_{ij} be defined by equation (2); let the rank of the matrix $||\tau_{ij}||$ defined by (3) be t .

(i) When $t = 0$, the joint distribution of the a_{ij} is given by the Wishart distribution.

(ii) When $t = 1$, the joint distribution of the a_{ij} is

$$(31) \quad \frac{e^{-\frac{1}{2} \sum_{i,j=1}^p \sum_{a=1}^N \sigma^{ij} (\mu_{ia} - \bar{\mu}_i) (\mu_{ja} - \bar{\mu}_j)}}{2^{\frac{1}{2}p(N-1) - \frac{1}{2}(N-3)} \pi^{\frac{1}{2}p(p-1)} \sum_{i=1}^{p-1} \Gamma(\frac{1}{2}[N-1-i])} | \sigma^{ij} |^{\frac{1}{2}(N-1)} | a_{ij} |^{\frac{1}{2}(N-p-2)}} \\ \times e^{\frac{1}{2} \sum_{i,j=1}^k \sigma^{ij} a_{ij}} \left[\sum_{i,j=1}^p \sum_{a=1}^N a_{ij} (\mu_{ia} - \bar{\mu}_i) (\mu_{ja} - \bar{\mu}_j) \right]^{-\frac{1}{2}(N-3)} \\ \times I_{\frac{1}{2}(N-3)} \left(\sqrt{\sum_{i,j=1}^p \sum_{a=1}^N a_{ij} (\mu_{ia} - \bar{\mu}_i) (\mu_{ja} - \bar{\mu}_j)} \right).$$

(iii) When $t = 2$, the joint distribution of the a_{ij} is given by

$$(32) \quad \frac{e^{-\frac{1}{2} \sum_{i,j=1}^p \sum_{a=1}^N \sigma^{ij} (\mu_{ia} - \bar{\mu}_i) (\mu_{ja} - \bar{\mu}_j)}}{2^{\frac{1}{2}p(N-1) - \frac{1}{2}(N-3)} \pi^{\frac{1}{2}p(p-1)} \prod_{i=1}^{p-2} \Gamma(\frac{1}{2}[N-2-i])} | \sigma^{ij} |^{\frac{1}{2}(N-1)} | a_{ij} |^{\frac{1}{2}(N-p-2)}} \\ \times e^{-\frac{1}{2} \sum_{i,j=1}^p \sigma^{ij} a_{ij}} \times \sum_{\omega=0}^{\infty} \frac{(u_1 u_2)^{\omega}}{2^{2\omega} \omega! \Gamma\left(\frac{N-2}{2} + \omega\right)} (u_1 + u_2)^{-\frac{1}{2}(N-3) + 2\omega} \\ \times I_{\frac{1}{2}(N-3) + 2\omega} (\sqrt{u_1 + u_2}),$$

where u_1 and u_2 are the two non-zero roots of (29).

(iv) When $t > 2$, the joint distribution of the a_{ij} can be written by means of expression (14) as a multiple integral. The explicit form of the distribution has not yet been obtained.

⁶ One method of demonstrating this fact is to apply (8) to centrally distributed variates and compare the Wishart distribution of the transformed variates with the Wishart distribution of the original variates.

REFERENCES

- [1] R. A. FISHER, "The statistical utilization of multiple measurements," *Ann. of Eug.*, Vol. 8 (1938), pp. 376-386
- [2] W. G. COCHRAN, "The comparison of different scales of measurements for experimental results," *Annals of Math Stat*, Vol. 14 (1943), pp. 205-216.
- [3] G. TINTNER, "An application of the variate difference method to multiple regression," *Econometrica*, Vol. 12 (1944), pp. 97-113.
- [4] H. HOTELLING, "Relations between two sets of variates," *Biometrika*, Vol. 28 (1936), pp. 336-337
- [5] M. BÔCHER, *Introduction to Higher Algebra* (1907), MacMillan Company, New York, pp. 163-170.
- [6] W. G. MADON, "Contributions to the theory of multivariate statistical analysis," *Trans of the Amer Math Soc*, Vol. 44 (1938), p. 476
- [7] E. GOURSAT (E. R. HEDRICK), *A Course in Mathematical Analysis*, (1904), Ginn and Company, Cambridge, Mass, p. 308.
- [8] G. N. WATSON, *A Treatise on the Theory of Bessel Functions*, (1944), MacMillan Company, New York.

STATISTICAL TESTS BASED ON PERMUTATIONS OF THE OBSERVATIONS

A. WALD AND J. WOLFOWITZ

Columbia University

1. Introduction. One of the problems of statistical inference is to devise exact tests of significance when the form of the underlying probability distribution is unknown. The idea of a general method of dealing with this problem originated with R. A. Fisher [13, 14]. The essential feature of this method is that a certain set of permutations of the observations is considered, having the property that each permutation is equally likely under the hypothesis to be tested. Thus, an exact test on the level of significance α can be constructed by choosing a proportion α of the permutations as critical region. In an interesting paper H. Scheffé [2] has shown that for a general class of problems this is the only possible method of constructing exact tests of significance.

Tests based on permutations of the observations have been proposed and studied by R. A. Fisher, E. J. G. Pitman, B. L. Welch, the present authors, and others. Pitman and Welch derived the first few moments of the statistics used in their test procedures. However, it is desirable to derive at least the limiting distributions of these statistics and make it practicable to carry out tests of significance when the sample is large. Such a large sample distribution was derived for a statistic considered elsewhere [1] by the present authors.

In this paper a general theorem on the limiting distribution of linear forms in the universe of permutations of the observations is derived. As an application of this theorem, the limiting distributions of the rank correlation coefficient and that of several statistics considered by Pitman and Welch, are obtained. In the last section the limiting distribution of Hotelling's generalized T in the universe of permutations of the observations is derived.

2. A theorem on linear forms. Let $H_N = (h_1, h_2, \dots, h_N)$ ($N = 1, 2, \dots$, ad inf.) be sequences of real numbers and let

$$\mu_r(H_N) = N^{-1} \sum_{\alpha=1}^N \left(h_{\alpha} - N^{-1} \sum_{\beta=1}^N h_{\beta} \right)^r$$

for all integral values of r . We define the following symbols in the customary manner: For any function $f(N)$ and any positive function $\varphi(N)$ let $f(N) = O(\varphi(N))$ mean that $|f(N)/\varphi(N)|$ is bounded from above for all N and let

$$f(N) = \Omega(\varphi(N))$$

mean that

$$f(N) = O(\varphi(N))$$

and that

$$\liminf_N |f(N)/\varphi(N)| > 0.$$

Also let

$$f(N) = o(\varphi(N))$$

mean that

$$\lim_{N \rightarrow \infty} f(N)/\varphi(N) = 0.$$

Let $[\rho]$ denote the largest integer $\leq \rho$

We shall say that the sequences H_N ($N = 1, 2, \dots$, ad inf.) satisfy the condition W if, for all integral $r > 2$,

$$(2.1) \quad \frac{\mu_r(H_N)}{[\mu_2(H_N)]^{r/2}} = O(1).$$

For any value of N let

$$X = (x_1, x_2, \dots, x_N)$$

be a chance variable whose domain of definition is made up of the $N!$ permutations of the elements of the sequence $A_N = (a_1, a_2, \dots, a_N)$. (If two of the a_i ($i = 1, 2, \dots, N$) are identical we assume that some distinguishing index is attached to each so that they can then be regarded as distinct and so that there still are $N!$ permutations of the elements a_1, \dots, a_N). Let each permutation of A_N have the same probability $(N!)^{-1}$. Let $E(Y)$ and $\sigma^2(Y)$ denote, respectively, the expectation and variance of any chance variable Y .

We now prove the following:

THEOREM *Let the sequences $A_N = (a_1, a_2, \dots, a_N)$ and $D_N = (d_1, d_2, \dots, d_N)$ ($N = 1, 2, \dots$, ad inf.) satisfy the condition W . Let the chance variable L_N be defined as*

$$L_N = \sum_{i=1}^N d_i x_i.$$

Then as $N \rightarrow \infty$, the probability of the inequality

$$L_N - E(L_N) < t \sigma(L_N)$$

for any real t , approaches

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}x^2} dx.$$

For convenience the proof will be divided into several lemmas.

Since

$$L_N^* = \frac{L_N - E(L_N)}{\sigma(L_N)}$$

remains invariant if a constant is added to all the elements of D_N or of A_N , or if the elements of either of the latter are multiplied by any constant other than zero, we may, in the formation of L_N^* , replace A_N and D_N by the sequences A'_N

and D'_N , respectively, whose i th elements a'_i and d'_i ($i = 1, 2, \dots, N$) are, respectively

$$(2.2) \quad a'_i = [\mu_2(A_N)]^{-1} \left(a_i - N^{-1} \sum_{j=1}^N a_j \right)$$

and

$$(2.3) \quad d'_i = [\mu_2(D_N)]^{-1} \left(d_i - N^{-1} \sum_{j=1}^N d_j \right).$$

The sequences A'_N and D'_N satisfy the condition W . Furthermore,

$$(2.4) \quad \mu_1(A'_N) \equiv \mu_1(D'_N) \equiv 0$$

and

$$(2.5) \quad \mu_2(A'_N) \equiv \mu_2(D'_N) \equiv 1.$$

LEMMA 1.

$$(2.6) \quad \sum_{\alpha_1 < \alpha_2 < \dots < \alpha_k \leq N} a'_{\alpha_1} a'_{\alpha_2} \dots a'_{\alpha_k} = O(N^{[k/2]})$$

$$(2.7) \quad \sum_{\alpha_1 < \alpha_2 < \dots < \alpha_k \leq N} d'_{\alpha_1} d'_{\alpha_2} \dots d'_{\alpha_k} = O(N^{[k/2]}).$$

From (2.4), (2.5), and the fact that the A'_N and D'_N satisfy condition W , it follows that the A'_N and D'_N satisfy conditions $a)$, $b)$, and $c)$ of the theorem on page 383 of [1]. Our lemma 1 is the same as lemma 1 of [1].

LEMMA 2. Let

$$V = (v_1, v_2, \dots, v_N)$$

be the same permutation of the elements of A'_N that X is of the elements of A_N . Let $y = v_1 \dots v_k z$ where $z = v'_{i_1} \dots v'_{i_r}$, $i_j > 1$ ($j = 1, 2, \dots, r$), and k, r, i_1, \dots, i_r are fixed values independent of N .

Then

$$(2.8) \quad E(y) = O(N^{[k/2]-k}).$$

This is Lemma 2 of [1].

In a similar manner we obtain that

$$(2.9) \quad \sum_{\alpha_1, \alpha_2, \dots, \alpha_{k+r}} d'_{\alpha_1} \dots d'_{\alpha_k} d'_{i_1} \dots d'_{i_r} = O(N^{[k/2]-k}) \cdot O(N^{k+r}) = O(N^{[k/2]+r}).$$

The summation in the above formula is to be taken over all possible sets of $k+r$ distinct positive integers $\leq N$.

LEMMA 3. Let $\alpha_1, \dots, \alpha_{k+r}$ be $(k+r)$ distinct positive integers $\leq N$. Then

$$(2.10) \quad E(v_1 v_2 \dots v_k v'_{i_1} \dots v'_{i_r}) = E(v_{\alpha_1} v_{\alpha_2} \dots v_{\alpha_k} v'_{i_1} \dots v'_{i_r}).$$

This follows from the fact that all permutations of A'_N have the same probability.

LEMMA 4. *Let*

$$L'_N = \sum_{i=1}^N d'_i v_i.$$

Then

$$(2.11) \quad E(L_N'^p) = O(N)^{[p/2]}.$$

PROOF: Expand $L_N'^p$ and take the expected value of the individual terms. The contribution to $E(L_N'^p)$ of all the terms which are multiples of the type appearing in the right member of (2.10) with fixed $k, r, i_1, \dots, i_r (k + i_1 + \dots + i_r = p)$, is, by Lemmas 2 and 3

$$\begin{aligned} O(N^{[k/2]-k}) \cdot \sum_{\substack{\alpha_1, \dots, \alpha_r (k+r) \\ \text{all different}}} \dots \sum d'_{\alpha_1} \dots d'_{\alpha_k} d'_{\alpha_{k+1}} \dots d'_{\alpha_{k+r}} &= O(N^{[k/2]-k}) O(N^{[k/2]+r}) \\ &= O(N^{2[k/2]-k+r}). \end{aligned}$$

Since $i_j > 1 (j = 1, \dots, r)$, it follows from the fact that $k + i_1 + \dots + i_r = p$ that $2r \leq p - k$ and that $2r = p - k$ only if $i_1 = \dots = i_r = 2$. Now

$$(2.12) \quad 2 \left[\frac{k}{2} \right] - k + r \leq r \leq \frac{p-k}{2} \leq \frac{p}{2}.$$

Hence the maximum value of $2 \left[\frac{k}{2} \right] - k + r$ is reached when $r = \left[\frac{p}{2} \right]$ and $k = 0$.

This proves the lemma.

From the last remarks of the preceding paragraph we obtain

LEMMA 5.

$$(2.13) \quad E(L_N'^{2j}) - \frac{(2j)!}{j!2^j} \left(\sum_{\substack{\alpha_1, \dots, \alpha_j \\ \text{all different}}} d_{\alpha_1}' \dots d_{\alpha_j}' \right) E(v_1^2 \dots v_j^2) = o(N^j).$$

We now prove

LEMMA 6.

$$(2.14) \quad E(L_N') = 0$$

$$(2.15) \quad E(L_N'^2) = NE(v_1^2) + o(N) = N + o(N).$$

Equation (2.14) follows from (2.2). Consider the expectations of the various terms in the expansion of $L_N'^2$. The sum of all the terms of the type

$$d_i' d_j' E(v_i v_j)$$

is

$$\left(\sum_{i \neq j} d_i' d_j' \right) E(v_i v_j) = O(N) O(N^{-1}) = O(1),$$

by Lemmas 1 and 2. The sum of all the terms of the type

$$d_i'^2 E(v_i^2)$$

is

$$\left(\sum_{i=1}^N d_i'^2 \right) E(v_1^2) = NE(v_1^2) = N,$$

by (2.2) and (2.3). This proves the lemma.

LEMMA 7

$$(2.16) \quad E(v_1^2 \cdots v_1^2) = 1 + o(1)$$

$$(2.17) \quad \sum \cdots \sum_{\substack{\alpha_1, \dots, \alpha_s \\ \text{all different}}} d_{\alpha_1}'^2 \cdots d_{\alpha_s}'^2 = N^s + o(N^s).$$

From (2.2) and (2.3), and Lemma 3, it follows that it will be sufficient to prove (2.17), because (2.16) follows in the same manner. Consider the relation

$$N^s = \left(\sum_{i=1}^N d_i'^2 \right)^s = \sum_{\substack{\alpha_1, \dots, \alpha_s \\ \text{all different}}} d_{\alpha_1}'^2 \cdots d_{\alpha_s}'^2 + \text{other terms}.$$

By (2.9) the sum of these other terms must be not larger than $O(N^{s-1})$. From this follows the lemma.

PROOF of the theorem: Since

$$L_N^* = \frac{L_N'}{\sigma(L_N')} = \frac{L_N - E(L_N)}{\sigma(L_N)},$$

it will be sufficient to show that the moments of L_N^* approach those of the normal distribution as $N \rightarrow \infty$. From (2.14), (2.15), and (2.11) we see that, when p is odd, the p th moment of L_N^* is $O(N^{-1/2})$ and hence approaches zero as $N \rightarrow \infty$. When p is even and $= 2s$ (say), it follows from Lemma 5 that

$$E(L_N'^{2s}) = \frac{(2s)!}{s!2^s} \left(\sum \cdots \sum_{\substack{\alpha_1, \dots, \alpha_s \\ \text{all different}}} d_{\alpha_1}'^2 \cdots d_{\alpha_s}'^2 \right) E(v_1^2 \cdots v_s^2) = o(N^s).$$

Hence from (2.16) and (2.17)

$$(2.18) \quad E(L_N'^{2s}) = \frac{(2s)!}{s!2^s} N^s + o(N^s).$$

From (2.18) and (2.15) we obtain that

$$\lim_{N \rightarrow \infty} E(L_N^{*2s}) = \frac{(2s)!}{s!2^s}.$$

This completes the proof of the theorem.

It will be noticed that nothing in the foregoing proof requires that, when $N < N'$, the sequences A_N and D_N be subsequences of $A_{N'}$ and $D_{N'}$. Indeed, the sequences were written as they were simply for typographic brevity. We have therefore

COROLLARY 1. *The theorem is valid for sequences*

$$A_N = (a_{N1}, \dots, a_{NN})$$

$$D_N = (d_{N1}, \dots, d_{NN})$$

$$(N = 1, 2, \dots \text{ ad inf.})$$

provided they fulfill condition W.

COROLLARY 2. *If the elements $a_i (i = 1, 2, \dots \text{ ad inf.})$ are all independent observations on the same chance variable, all of whose moments are finite and whose variance is positive, the sequences $A_N (N = 1, 2, \dots \text{ ad inf.})$ will fulfill condition W with probability one*

3. The rank correlation coefficient. For this well known statistic (see [3])

$$A_N \equiv D_N \equiv (1, 2, 3, \dots, N).$$

The sequences A_N and D_N satisfy the condition W. For

$$\sum_{i=1}^N i^r = O(N^{r+1})$$

and hence, for $r \geq 3$

$$\mu_r(A_N) = \mu_r(D_N) = O(N^r).$$

Also

$$\mu_2(A_N) = \mu_2(D_N) = \Omega(N^2).$$

Hence the distribution of the rank correlation coefficient is asymptotically normal in the case of statistical independence. This result was first proved by Hotelling and Pabst [3].

4. Pitman's test for dependence between two variates. The distribution of the correlation coefficient in the population of permutations of the observations was used by Pitman [4] in a test for dependence between two variates which "involves no assumptions" about the distributions of these variates. In our notation, let

$$(a_i, d_i) (i = 1, 2, \dots, N)$$

be N observations on the pair of variates A and D whose dependence it is desired to test. Then the value of the correlation coefficient is

$$N^{-1} \sum_{i=1}^N d'_i a'_i.$$

At the level β the observations are considered to be significant if the probability that $N^{-1} |L'_N|$ be equal to or greater than the absolute value of the actually observed correlation coefficient is $\leq \beta$.

In his paper ([4], page 227) Pitman points out that if the ratios of certain sample cumulants are "not too large," then, as $N \rightarrow \infty$, the first four moments of $N^{-1} L'_N$ will approach 0, 1, 0, and 3, respectively (the first moment is always zero). Our theorem and the relation (2.15) make clear that under proper circumstances all the moments will approach those of the normal distribution.

5. Pitman's procedure for testing the hypothesis that two samples are from the same population. For testing the hypothesis that two samples came from the same population Pitman [5] proposed the following procedure:
Let one sample be

$$a_1, a_2, \dots, a_m$$

and the other

$$a_{m+1}, a_{m+2}, \dots, a_{m+n}.$$

Write $m + n = N$, and construct the sequences A_N and A'_N as before defined. Let

$$\begin{aligned} d_i &= 1 & (i = 1, \dots, m) \\ d_i &= 0 & (i = m+1, \dots, N) \end{aligned}$$

and construct the sequences D_N and D'_N . Then the value of the statistic considered by Pitman is, except for a constant factor,

$$(5.1) \quad N^{-1} \left(\sum_{i=1}^N d'_i a'_i \right).$$

At the level β the observations are considered significant if the probability that $N^{-1} |L'_N|$ be equal to or greater than the observed absolute value of the expression (5.1) is $\leq \beta$.

Let $N \rightarrow \infty$, while $\frac{m}{n}$ is constant. Then the sequences D_N are seen to satisfy condition *W*. If then the sequences A_N satisfy condition *IV* we may, for large N , employ the result of our theorem and expeditiously determine the critical value of Pitman's statistic.

6. Analysis of variance in randomized blocks. Welch [7] and Pitman [6] consider the following problem: Each of n different "varieties of a plant" is planted in one of the n cells which constitute a "block." It is desired to test, on the basis of results from m blocks, the null hypothesis that there is no difference among the varieties. In order to eliminate a possible bias caused by variations in fertility among the cells of a block, the varieties are assigned at random to the cells of a block. If the cells of the j th block are designated by $(j1), (j2), \dots, (jn)$, a permutation of the integers $1, 2, \dots, n$ is allocated to the j th block by a chance process, each permutation having the same probability $(n!)^{-1}$.

Let x_{ijk} be the yield of the i th variety in the k th cell of the j th block to which it was assigned by the randomization process. It is assumed that

$$x_{i,k} = \mu_{i,k} + \delta_i + \epsilon_{i,k},$$

where y_{jk} is the "effect" of the k th cell in the j th block, δ_i is the "effect" of the i th variety, and ϵ_{jk} are chance variables about whose distribution we assume nothing. The null hypothesis states that

$$\delta_1 = \delta_2 = \dots = \delta_n = 0.$$

Let a_{jk} be the yield in the k th cell of the j th block and x_{ij} the yield of the i th variety in the j th block. If the null hypothesis is true then, because of the randomization within each block described above, the conditional probability that, given the set $\{a_{jk}\} (k = 1, 2, \dots, n)$, the sequence $x_{1j}, x_{2j}, \dots, x_{nj}$, be any given permutation of the elements of $\{a_{jk}\}$ is $(n!)^{-1}$. Permuting in all the blocks simultaneously we have that, under the null hypothesis, given the set of mn values $\{a_{jk}\} (j = 1, 2, \dots, m; k = 1, 2, \dots, n)$, the conditional probability of any of the permutations is the same, $(n!)^{-m}$. This permits an exact test of the null hypothesis.

The classical analysis of variance statistic that would be employed in the conventional two-way classification with independent normally distributed observations is

$$F = \frac{(m-1)m \sum (x_{i.} - \bar{x})^2}{\sum \sum (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2}$$

where

$$\begin{aligned} x_{i.} &= m^{-1} \sum_j x_{ij} \\ x_{.j} &= n^{-1} \sum_i x_{ij} \end{aligned} \quad x = (mn)^{-1} \sum \sum x_{ij}.$$

The statistic W used by Welch and Pitman is

$$W = F(m-1+F)^{-1}.$$

Since W is a monotonic function of F and the critical regions are the upper tails, the two tests are equivalent. The distribution of F or W is to be determined in the same manner as that of the other statistics discussed in this paper, i.e., over the equally probable permutations of the values actually observed. The critical region is, as usual, the upper tail.

Since $x_{i.}$ takes any of the values a_{j1}, \dots, a_{jn} with probability $1/n$, we have

$$(6.1) \quad E(x_{i.}) = n^{-1} \sum_k a_{jk} = a_j \quad (\text{say}).$$

$$(6.2) \quad \sigma^2(x_{i.}) = n^{-1} \sum_k (a_{jk} - a_j)^2 = b_j \quad (\text{say}).$$

$$\begin{aligned} \sigma(x_{i_1j} x_{i_2j}) &= [n(n-1)]^{-1} \sum_{k_1 \neq k_2} a_{jk_1} a_{jk_2} - a_j^2 \\ &= [n(n-1)]^{-1} [(\sum_k a_{jk}^2) - \sum_k a_{jk}^2] - a_j^2 \\ (6.3) \quad &= [n^2 a_j^2 - \sum_k a_{jk}^2] [n(n-1)]^{-1} - a_j^2 \\ &= (n-1)^{-1} [a_j^2 - n^{-1} \sum_k a_{jk}^2] = -(n-1)^{-1} b_j. \end{aligned}$$

Hence

$$(6.4) \quad E(x_{i.}) = m^{-1} \sum a_j.$$

$$(6.5) \quad \sigma^2(x_{i.}) = m^{-2} \sum b_j = b \quad (\text{say}).$$

$$(6.6) \quad \sigma(x_{i_1.}, x_{i_2.}) = -[m^2(n-1)]^{-1} \sum b_j = c \quad (\text{say}).$$

$$i_1 \neq i_2$$

Let

$$x_{ij}^* = \sum_v \lambda_{iv} x_{vj} \quad (i, v = 1, \dots, n)$$

where $\|\lambda_{iv}\|$ is an orthogonal matrix and

$$\lambda_{n1} = \lambda_{n2} = \dots = \lambda_{nn} = n^{-1/2}.$$

Then it follows that

$$(6.7) \quad \begin{aligned} E(x_{i.}^*) &= 0 \\ \sigma^2(x_{i.}^*) &= b - c \quad (i = 1, 2, \dots, n-1) \\ \sigma(x_{i_1.}^*, x_{i_2.}^*) &= 0 \quad (i_1 \neq i_2, i_1, i_2 = 1, \dots, n-1). \end{aligned}$$

Furthermore, we have

$$(6.8) \quad \sum_{i=1}^{n-1} x_{i.}^{*2} = \sum_{i=1}^n (x_{i.} - x)^2.$$

Applying the well known identity

$$\Sigma \Sigma (x_{ij} - x_{i.} - x_{.j} + x)^2 = \Sigma \Sigma (x_{ij} - x_{.j})^2 - m \Sigma (x_{i.} - x)^2$$

to the definitions of F and W we obtain

$$(6.9) \quad W = \frac{m \sum_i (x_{i.} - x)^2}{\sum_i \sum_j (x_{ij} - x_{.j})^2}.$$

The denominator of the right member of (6.9) is invariant under permutations *within* each block and equals

$$\sum_j \sum_k (a_{jk} - a_j)^2 = (n-1)m^2(b-c).$$

Hence

$$(6.10) \quad \begin{aligned} W &= [m(n-1)(b-c)]^{-1} \sum_{i=1}^n (x_{i.} - x)^2 \\ &= [m(n-1)(b-c)]^{-1} \sum_{i=1}^{n-1} x_{i.}^{*2}. \end{aligned}$$

If the joint distribution of the x_i^* ($i = 1, 2, \dots, n-1$) over the set of admissible permutations approaches a normal distribution with non-singular correlation

matrix as m , the number of blocks, becomes large, it follows from (6.7) and (6.10) that the distribution of $m(n-1)W$ approaches the χ^2 distribution with $n-1$ degrees of freedom. Hence it remains to indicate conditions on the set $\{a_{jk}\}$ which would make the distribution of the x_i^* approach normality. Each x_i^* is the mean of independent variables, so these conditions need not be very restrictive.

According to Cramér [8], Theorem 21a, page 113, if the variances and covariances fulfill certain requirements (the limiting correlation matrix should also be non-singular) and if a generalized Lindeberg condition holds, normality in the limit will follow. Somewhat more restrictive conditions which are simpler to state and which will be satisfied in most statistical applications are that $a < c' < b_j < c''$ for all j , where c' and c'' are positive constants. Since the variance of x_i^* is $(n-1)^{-1}nb_j$, it can be seen that the above inequalities imply the fulfillment of the conditions of the Laplace-Liapounoff theorem (see, for example, Uspensky [9], page 318). By [6.7] the correlation matrix is non-singular.

7. Hotelling's generalized T for permutation of the observations. In this section we shall restrict ourselves to bivariate populations, the extension to more than two variables being straightforward. Let $(u_{11}, u_{21}), \dots, (u_{1m}, u_{2m})$ be m pairs of observations on the chance variables U_1, U_2 , and $(u_{1(m+1)}, u_{2(m+1)}), \dots, (u_{1N}, u_{2N})$, be n pairs of observations on the chance variables U'_1, U'_2 , where $m+n=N$. If each of the pairs U_1, U_2 , and U'_1, U'_2 is jointly normally distributed with the same covariance matrix, the Hotelling generalized T for testing the null hypothesis that

$$(7.1) \quad E(U_1) = E(U'_1)$$

and

$$(7.2) \quad E(U_2) = E(U'_2),$$

is given (Hotelling [10]) by

$$T^2 = N^{-1}(mn) \sum_{j=1}^2 \sum_{i=1}^2 q_{ij}(\bar{u}_i - \bar{u}'_i)(\bar{u}_j - \bar{u}'_j)$$

where

$$m\bar{u}_i = \sum_{l=1}^m u_{li}, \quad n\bar{u}'_i = \sum_{l=m+1}^N u_{li}$$

and the matrix $\|q_{ij}\|$ is the inverse of the matrix $\|b_{ij}\|$ with b_{ij} given by

$$(N-2)b_{ij} = \sum_{l=1}^m (u_{li} - \bar{u}_i)(u_{lj} - \bar{u}_j) + \sum_{l=m+1}^N (u_{li} - \bar{u}'_i)(u_{lj} - \bar{u}'_j).$$

In Hotelling's procedure the b_{ij} are sample estimates of the population covariances whose distribution is independent of that of the sample means. A constant multiple of the statistic T^2 has the analysis of variance distribution under the null hypothesis. If the population covariances were known and used in place of the b_{ij} , T^2 would have the χ^2 distribution with two degrees of freedom.

Let us now apply the generalized T over the permutations of the actually observed values, as was done with other statistics in previous sections. If we do this literally we will find that the b_i , are no longer independent of the sample means. To avoid this complication we shall use a slightly different statistic T' which, as will be shown later, is a monotonic function of T , so that the test based on T' is identical with that based on T . The statistic T' is defined as follows: Let

$$\bar{U}_i = N^{-1} \sum_{k=1}^N u_{ik}$$

$$c'_{ij} = N[(N-1)mn]^{-1} \sum_{k=1}^N (u_{ik} - \bar{U}_i)(u_{jk} - \bar{U}_j) \quad (i, j = 1, 2)$$

and

$$\|q'_{ij}\| = \|c'_{ij}\|^{-1}.$$

Then

$$(7.3) \quad T'^2 = \sum_{i=1}^2 \sum_{j=1}^2 q'_{ij}(\bar{u}_i - \bar{u}'_i)(\bar{u}_j - \bar{u}'_j).$$

The expression T'^2 is much simpler than T^2 , since the coefficients q'_{ij} are constants in the population of permutations of the observations. We shall show that T'^2 is a monotonic function of T^2 . Let

$$Q_{ij} = \sum_{k=1}^m (u_{ik} - \bar{u}_i)(u_{jk} - \bar{u}_j) + \sum_{k=m+1}^N (u_{ik} - \bar{u}'_i)(u_{jk} - \bar{u}'_j)$$

$$Q'_{ij} = \sum_{k=1}^N (u_{ik} - \bar{U}_i)(u_{jk} - \bar{U}_j)$$

$$\|Q^{ij}\| = \|Q_{ij}\|^{-1}$$

$$\|Q'^{ij}\| = \|Q'_{ij}\|^{-1}.$$

Then the expressions

$$(7.4) \quad T_1^2 = \sum_{i=1}^2 \sum_{j=1}^2 Q^{ij}(\bar{u}_i - \bar{u}'_i)(\bar{u}_j - \bar{u}'_j)$$

and

$$(7.5) \quad T_2^2 = \sum_{i=1}^2 \sum_{j=1}^2 Q'^{ij}(\bar{u}_i - \bar{u}'_i)(\bar{u}_j - \bar{u}'_j),$$

are constant multiples of T^2 and T'^2 , respectively. Hence it is sufficient to show that T_2^2 is a monotonic function of T_1^2 . We have

$$(7.6) \quad Q'_{ij} = Q_{ij} + m(\bar{u}_i - \bar{U}_i)(\bar{u}_j - \bar{U}_j) + n(\bar{u}'_i - \bar{U}_i)(\bar{u}'_j - \bar{U}_j).$$

Furthermore, we have

$$(7.7) \quad \bar{u}_i - \bar{U}_i = \bar{u}_i - \frac{m\bar{u}_i + n\bar{u}'_i}{m+n} = \frac{n(\bar{u}_i - \bar{u}'_i)}{m+n}.$$

Similarly

$$(7.8) \quad \bar{u}'_i - \bar{U}_i = \bar{u}'_i - \frac{m\bar{u}_i + n\bar{u}'_i}{m+n} = \frac{m(\bar{u}'_i - \bar{u}_i)}{m+n}.$$

From (7.6), (7.7) and (7.8) it follows that

$$(7.9) \quad \begin{aligned} Q'_{ij} &= Q_{ij} + \frac{mn^2}{(m+n)^2} (\bar{u}_i - \bar{u}'_i)(\bar{u}_j - \bar{u}'_j) + \frac{nm^2}{(m+n)^2} (\bar{u}_i - \bar{u}'_i)(\bar{u}_j - \bar{u}'_j) \\ &= Q_{ij} + \frac{mn}{m+n} (\bar{u}_i - \bar{u}'_i)(\bar{u}_j - \bar{u}'_j). \end{aligned}$$

Denote $\frac{mn}{m+n}$ by λ and $\bar{u}_i - \bar{u}'_i$ by h_i . Then we have

$$(7.10) \quad Q'_{ij} = Q_{ij} + \lambda h_i h_j.$$

Denote the cofactor of Q_{ij} in $||Q_{ij}||$ by R_{ij} , and the cofactor of Q'_{ij} in $||Q'_{ij}||$ by R'_{ij} . Then

$$(7.11) \quad \frac{|Q_{ij}|}{|Q'_{ij}|} = \frac{|Q_{ij}|}{|Q_{ij} + \lambda h_i h_j|} = \frac{|Q_{ij}|}{|Q_{ij}| + \lambda \Sigma \Sigma R_{ij} h_i h_j} = \frac{1}{1 + \lambda T_1^2}.$$

Furthermore, we have

$$(7.12) \quad \frac{|Q_{ij}|}{|Q'_{ij}|} = \frac{|Q'_{ij} - \lambda h_i h_j|}{|Q'_{ij}|} = \frac{|Q'_{ij}| - \lambda \Sigma \Sigma R'_{ij} h_i h_j}{|Q'_{ij}|} = 1 - \lambda T_2^2.$$

From (7.11) and (7.12) it follows that T_2^2 is a monotonic function of T_1^2 . Hence also T'^2 is a monotonic function of T^2 and, therefore, we do not change our test procedure by using T'^2 instead of T^2 .

Let the sequence of pairs

$$(x_{11}, x_{21}), \dots, (x_{1N}, x_{2N})$$

be a permutation of the actually observed pairs

$$(u_{11}, u_{21}), \dots, (u_{1N}, u_{2N})$$

where to each permutation is ascribed the same probability $(N!)^{-1}$. Then one obtains for $i = 1, 2$,

$$(7.13) \quad E(\bar{x}_i - \bar{x}'_i) = 0$$

$$(7.14) \quad \sigma^2(\bar{x}_i - \bar{x}'_i) = N[(N-1)mn]^{-1} \sum_{j=1}^N (u_{ij} - \bar{U}_i)^2 = c'_{ii}$$

$$(7.15) \quad E(\bar{x}_1 - \bar{x}'_1)(\bar{x}_2 - \bar{x}'_2) = N[(N-1)mn]^{-1} \sum_{j=1}^N (u_{1j} - \bar{U}_1)(u_{2j} - \bar{U}_2) = c'_{12}.$$

Hence $||c'_i||$ is the covariance matrix of the variates

$$(\bar{x}_1 - \bar{x}'_1) \quad \text{and} \quad (\bar{x}_2 - \bar{x}'_2).$$

Now we shall show that the limiting distribution of T^{v^2} , as $N \rightarrow \infty$, is the χ^2 distribution with 2 degrees of freedom, provided that the observation u_{ik} ($i = 1, 2; k = 1, \dots, N$) satisfy some slight restrictions. Since $\|q'_{ik}\|$ is the inverse of the covariance matrix $\|c'_{ik}\|$, our statement about the limiting distribution of T^{v^2} is obviously proved if we can show that $\bar{x}_1 - \bar{x}'_1$ and $\bar{x}_2 - \bar{x}'_2$ have a joint normal distribution in the limit.

Let $N \rightarrow \infty$ while m/n remains constant. Let the sequences A_N and D_N of Section II be defined as follows:

There are two sequences A_N , denoted respectively by A_{1N} and A_{2N} , such that

$$a_{ij} = u_{ij}, \quad (i = 1, 2; j = 1, \dots, N).$$

Also

$$d_j = \frac{1}{m} \quad (j = 1, \dots, m)$$

$$d_j = -\frac{1}{n} \quad (j = m+1, \dots, N).$$

Then the sequences D_N satisfy the condition W . If also the sequences A_{iN} satisfy the condition W , the distribution of $\bar{x}_1 - \bar{x}'_1$ approaches the normal distribution as N increases, by the theorem of Section 2. If the joint distribution of $\bar{x}_1 - \bar{x}'_1$ and $\bar{x}_2 - \bar{x}'_2$ approaches a normal distribution with non-singular correlation matrix, the distribution of T^{v^2} approaches that of χ^2 with two degrees of freedom.

The correlation matrix of $(\bar{x}_1 - \bar{x}'_1)$ and $(\bar{x}_2 - \bar{x}'_2)$ will be of rank two in the limit if the correlation coefficient between $(\bar{x}_1 - \bar{x}'_1)$ and $(\bar{x}_2 - \bar{x}'_2)$ approaches a limit ρ , where $|\rho| < 1$. By (7.14) and (7.15) this is equivalent to saying that the absolute value of the angle between the vectors A'_{1N} and A'_{2N} is eventually greater than a positive lower bound. We shall show that, if the correlation coefficient approaches, as $N \rightarrow \infty$, a limit ρ whose absolute value is less than one, and if A_{1N} and A_{2N} satisfy the condition W , then $(\bar{x}_1 - \bar{x}'_1)$ and $(\bar{x}_2 - \bar{x}'_2)$ are jointly normally distributed in the limit.

Let δ_1 and δ_2 be any two real numbers not both zero. Then the sequence

$$A_N^* = (a_1^*, \dots, a_N^*)$$

where

$$a_i^* = \delta_1 a_{i1} + \delta_2 a_{i2}$$

will be shown to satisfy the condition W . If either δ_1 or δ_2 is zero this is trivial; assume therefore that neither is zero. Without loss of generality we may assume

that $\sum_{i=1}^N a_{i1} = 0$, for if this were not so we could replace the original a_{ij} by $a'_{ij} = a_{ij} - N^{-1} \sum_j a_{ij}$ as was done in Section 2. Let ρ' be such that $1 > \rho' > |\rho|$.

For N sufficiently large we have

$$\begin{aligned}\mu_2(A_N^*) &\geq N^{-1}(\delta_1^2 \sum_j a_{1j}^2 - 2|\delta_1 \delta_2 \sum a_{1j} a_{2j}| + \delta_2^2 \sum_j a_{2j}^2) \\ &\geq N^{-1}(\delta_1^2 \sum_j a_{1j}^2 - 2\rho' |\delta_1 \delta_2| \sqrt{(\sum_j a_{1j}^2)(\sum_j a_{2j}^2)} + \delta_2^2 \sum_j a_{2j}^2) \\ &= N^{-1}[(|\delta_1| \sqrt{\sum_j a_{1j}^2} - |\delta_2| \sqrt{\sum_j a_{2j}^2})^2 \\ &\quad + 2(1 - \rho') |\delta_1 \delta_2| \sqrt{(\sum_j a_{1j}^2)(\sum_j a_{2j}^2)}]\end{aligned}$$

and

$$\mu_2(A_N^*) \leq 2(\delta_1^2 \mu_2(A_{1N}) + \delta_2^2 \mu_2(A_{2N})).$$

Hence

$$(7.16) \quad \mu_2(A_N^*) = \Omega[\max\{\mu_2(A_{1N}), \mu_2(A_{2N})\}]$$

Also $\mu_r(A_N^*)$ is a sum of constant multiples of terms of the type

$$N^{-1} \sum_j a_{1j}^r a_{2j}^{r-1}$$

By Schwarz' inequality

$$(7.17) \quad N^{-1} \sum_j a_{1j}^r a_{2j}^{r-1} \leq N^{-1}(\sum_j a_{1j}^{2r})^{\frac{1}{2}} (\sum_j a_{2j}^{2(r-1)})^{\frac{1}{2}} = (\mu_{2r}(A_{1N}) \mu_{2(r-1)}(A_{2N}))^{\frac{1}{2}}.$$

The required result follows from (7.16) and (7.17).

Since the sequences A_N^* satisfy the condition W , the limiting distribution of

$$\delta_1(\bar{x}_1 - \bar{x}'_1) + \delta_2(\bar{x}_2 - \bar{x}'_2),$$

for any pair δ_1, δ_2 not both zero, is normal. From this and a theorem of Cramér and Wold ([11] Theorem 1, see also [8], Theorem 31) it follows that if the joint distribution of $(\bar{x}_1 - \bar{x}'_1)$ and $(\bar{x}_2 - \bar{x}'_2)$ approaches a limit, this limit must be the normal distribution. From a theorem of Radon ([12], see also Cramér [8], page 101) it follows that if the joint distribution of $(\bar{x}_1 - \bar{x}'_1)$ and $(\bar{x}_2 - \bar{x}'_2)$ does not approach a limit as $N \rightarrow \infty$ it is possible to find two subsequences of the sequence $(1, 2, \dots, N, \dots)$ ad inf. for each of which the joint distribution approaches a different limit. This contradicts the previous result. Hence the limit exists and is the normal distribution. This proves our statement that the limiting distribution of T'^2 is the χ^2 distribution with two degrees of freedom.

The statistic T'^2 seems to be appropriate for testing the null hypothesis that two bivariate distributions Π_1 and Π_2 are identical if the alternatives are restricted to the case where Π_2 differs from Π_1 only in the mean values, i.e., the distribution Π_2 can be obtained from Π_1 by a translation. This is no restriction as compared with Hotelling's T -test since also the T -test is based on the assumption that the two normal populations differ at most in their mean values, i.e., the covariance matrices in the two populations are assumed to be equal.

REFERENCES

- [1] A. WALD AND J. WOLFOWITZ, *Annals of Math. Stat.*, Vol. 14 (1943), p. 378.
- [2] HENRY SCHEFFÉ, *Annals of Math. Stat.*, Vol. 14 (1943), p. 305.
- [3] H. HOTELLING AND M. R. PABST, *Annals of Math. Stat.*, Vol. 7 (1936), p. 29.
- [4] E. J. G. PITMAN, *Supp. Jour. Roy. Stat. Soc.*, Vol. 4 (1937), p. 225.
- [5] E. J. G. PITMAN, *Supp. Jour. Roy. Stat. Soc.*, Vol. 4 (1937), p. 119.
- [6] E. J. G. PITMAN, *Biometrika*, Vol. 29 (1938), p. 322.
- [7] B. L. WELCH, *Biometrika*, Vol. 29 (1937), p. 21.
- [8] HARALD CRAMÉR, *Random Variables and Probability Distributions*, Cambridge, 1937.
- [9] J. V. UBPENSKY, *Introduction to Mathematical Probability*, New York, 1937.
- [10] H. HOTELLING, *Annals Math. Stat.*, Vol. 2 (1931), p. 359.
- [11] H. CRAMÉR AND H. WOLD, *Journal London Math. Soc.*, Vol. 11 (1936), pp. 290-4.
- [12] J. RADON, *Sitzungsberichte Akademie Wien*, Vol. 122 (1913), pp. 1295-1438.
- [13] R. A. FISHER, *The Design of Experiments*, Edinburgh, 1937. (Especially Section 21.)
- [14] R. A. FISHER, *Statistical Methods for Research Workers*, Edinburgh, 1936 (Especially Section 21 02.)

ERROR CONTROL IN MATRIX CALCULATION

F. E. SATTERTHWAITE

Aetna Life Insurance Company

1. Introduction. The solutions of large sets of simultaneous equations and the inversion of matrices are often complicated by the fact that errors, such as those introduced by rounding, become magnified in the course of the calculations to such an extent that the results are useless. In this paper we shall show that if the norm of the matrix $A - I$ is less than 0.35, operations involving the inversion A or the multiplication by A^{-1} will be in a state of error control for "Doolittle" methods of calculation. Thus such calculations may be carried through with assurance that the errors in the results will be limited to two or three significant figures. We also point out that as soon as an approximation to A^{-1} is available, most problems may be restated to bring them within the requirements for error control. Therefore the solution can be immediately completed to the desired degree of accuracy in one step instead of requiring multiple steps as do the iterative methods.

2. The inversion of special matrices. Consider the problem of inverting the matrix $(I + F)$ where I is the identity matrix and $(I + F)$ is a non-singular square matrix. Let

$$(2.1) \quad G = (I + F)^{-1}.$$

Then

$$(2.2) \quad (I + F)G = I$$

or

$$(2.3) \quad G = I - FG.$$

In ordinary algebra this would not be a practical formula for the calculation of G . However in matrix algebra the situation may be different. Examine the expanded form of G :

$$(2.4) \quad g_{ij} = \delta_{ij} - \sum_k f_{ik} g_{kj}.$$

The summation is over all values of k from 1 to n . Next examine the affect of imposing certain restrictions on F . For example, let $f_{ij} = 0$ if $j \geq i$. This is equivalent to making the summation in (2.4) over the range 1 to $i - 1$. The first row of (2.4) then becomes

$$g_{1j} = \delta_{1j}$$

and no g 's appear on the right. For the second row

$$g_{2j} = \delta_{2j} - f_{21}g_{1j}.$$

The only g 's on the right are those on the first row which have already been calculated. For the third row

$$g_{31} = \delta_{31} - f_{31}g_{11} - f_{32}g_{21}.$$

The only g 's on the right are in the first and second rows and have already been calculated. Similarly for the fourth and later rows.

Thus it is seen that if F is a "pre-diagonal" matrix, (2.3) is a very simple and practical formula for the numerical calculation of the inverse of $(I + F)$. Also if F is a post-diagonal matrix, (2.3) may be used by working up from the bottom row.

Similarly, if a matrix H is to be multiplied by the inverse of $(I + F)$, let

$$(2.5) \quad G = (I + F)^{-1}H$$

and the working equation becomes

$$(2.6) \quad G = H - FG,$$

The inversion of a diagonal matrix is accomplished by inverting each of its diagonal elements. That is if

$$(2.7) \quad F = \|\delta_{ij}s_{ij}\|$$

then

$$(2.8) \quad F^{-1} = \|\delta_{ij}s_{ij}^{-1}\|.$$

3. The inversion of general matrices. The general inversion problem will be solved if a general matrix can be factored into matrices of the special types treated in the last section. For the moment assume that such a factorization is possible and let the factors of the general matrix, A , be

$$(3.1) \quad A = (R_1 + I)S_1(I + T_1)$$

where R_1 is a pre-diagonal matrix, S_1 a diagonal matrix, and T_1 a post-diagonal matrix. Then

$$(3.2) \quad A = S_1 + R_1S_1 + S_1T_1 + R_1S_1T_1.$$

A slight change in form now appears desirable so let

$$(3.3) \quad \begin{aligned} A &= (RS^{-1} + I)S(I + S^{-1}T) \\ &= R + S + T + RS^{-1}SS^{-1}T \end{aligned}$$

$$(3.4) \quad = R + S + T + RS^{-1}T.$$

For convenience let

$$(3.5) \quad B = R + S + T$$

and remember that R , S , and T have no non-zero elements in common. Therefore the non-zero elements of R , S , and T are equal to the corresponding elements of B . Rearranging (3.4) gives

$$(3.6) \quad B = A - RS^{-1}T$$

and the elements of B are determined by

$$(3.7) \quad b_{ij} = a_{ij} - \sum \frac{r_{ik} t_{kj}}{s_{kk}}.$$

Since $r_{ik} = 0$ for $k \geq i$, there is no point in making the summation beyond $k = i - 1$. Also since $t_{kj} = 0$ for $k \geq j$, there is no point in making the summation beyond $j - 1$. Therefore the summation in (3.7) is to be considered to be over the range 1 to the smaller of $i - 1$ and $j - 1$. The r 's, s 's and t 's on the right of (3.7) can now be replaced by the corresponding b 's:

$$(3.8) \quad b_{ij} = a_{ij} - \sum \frac{b_{ik} b_{kj}}{b_{kk}}.$$

Since the first row (column) of b 's equal the first row (column) of a 's, the second row (column) of b 's is a function of only those b 's in the first row and the first column, etc., any calculation routine which works down from the top and from the left to the right will lead to a ready determination of all the b 's by (3.8).

Thus we see that the assumed factorization (3.3) of A is always possible (unless some of the diagonal elements, b_{kk} , of B are zero) and moreover the elements of the factors are readily calculated by the simple equations, (3.8).

Therefore, to invert the general non-singular square matrix A , calculate the elements of an intermediate matrix $B = R + S + T$ by equations (3.5) and (3.7). Then from (3.3) we have

$$(3.9) \quad A^{-1} = (I + S^{-1}T)^{-1} S^{-1}(I + RS^{-1})^{-1}$$

which can be readily calculated by the methods of (2.3) and (2.6).

4. The Doolittle method. The Doolittle method of matrix calculation can now be expressed in terms of the matrices R , S , and T studied above. To illustrate we shall use the set of equations:

$$(4.1) \quad \begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= c_{11}y_1 + c_{12}y_2 + c_{13}y_3 = d_1, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= c_{21}y_1 + c_{22}y_2 + c_{23}y_3 = d_2, \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= c_{31}y_1 + c_{32}y_2 + c_{33}y_3 = d_3. \end{aligned}$$

This set of equations will be represented in the form of a three element matrix,

$$(4.1.a) \quad || AX \cdot CY \cdot D ||.$$

The essential feature of the Doolittle method of solution is that we replace (4.1) by an equivalent set of equations for which the prediagonal coefficients of the X 's are all zero and the diagonal coefficients are all unity. Therefore consider the set formed as follows:

$$(4.2) \quad \| A_3 X : C_3 Y : D_3 \| = S^{-1} \{ \| AX : CY : D \| - R \| A_3 Y : C_3 Y : D_3 \| \}.$$

Then

$$(4.3) \quad A_3 = S^{-1} \{ A - RA_3 \}$$

or

$$(4.4) \quad (S + R)A_3 = A$$

or

$$(4.5) \quad A_3 = (S + R)^{-1} (RS^{-1} + I) S (I + S^{-1}T) \quad \text{by (3.3)} \\ = (I + S^{-1}T).$$

Since $S^{-1}T$ is a postdiagonal matrix, $\| A_3 X : C_3 Y : D_3 \|$ are the required intermediate equations for a Doolittle type of solution.

The final solution is now easily obtained. Consider

$$(4.6) \quad \| A_4 X : C_4 Y : D_4 \| = \| A_3 X : C_3 Y : D_3 \| - (S^{-1}T) \| A_4 X : C_4 Y : D_4 \|.$$

We have

$$(4.7) \quad A_4 = A_3 - (S^{-1}T)A_4$$

or

$$(4.8) \quad (I + S^{-1}T)A_4 = A_3 \\ = I + S^{-1}T \quad \text{by (4.5).}$$

Therefore A_4 is in fact the identity matrix and (4.6) can be rewritten

$$(4.9) \quad \| X : C_4 Y : D_4 \| = \| A_3 X : C_3 Y : D_3 \| - (S^{-1}T) \| X : C_4 Y : D_4 \|$$

5. The non-symmetric case. In actual practice, the work has to be so arranged that the elements of the matrices R , S , and $(S^{-1}T)$ are set out so as to be readily available for use as multipliers in forming the intermediate and final sets of equations. Table I gives such a practical layout for the non-symmetric case.

The elements of $(S^{-1}T)$ are set out as the postdiagonal elements of A_3 so that they do not need further attention. To determine the elements of R and S , we form a set of pre-intermediate equations:

$$(5.1) \quad \| BX : \dots : \dots \| = \| AX : \dots : \dots \| - R \| A_3 X - X : \dots : \dots \|.$$

TABLE I

Layout of Doolittle solution for the non-symmetric case Coefficients not used further are indicated by ...

$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = c_{11}y_1 + c_{12}y_2 + c_{13}y_3 = d_1$	
$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = c_{21}y_1 + c_{22}y_2 + c_{23}y_3 = d_2$	
$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = c_{31}y_1 + c_{32}y_2 + c_{33}y_3 = d_3$	
$s_1x_1 + \dots = \dots = d_{21} = d_1$	
$x_1 + a_{12}x_2 + a_{13}x_3 = c_{31}y_1 + c_{32}y_2 + c_{33}y_3 = d_{31} = d_1/s_1$	
$r_{21}x_1 + s_2x_2 = \dots = d_{22} = d_2 - r_{21}[d_{31} - x_1]$	
$x_2 + a_{22}x_2 + a_{23}x_3 = c_{32}y_2 + c_{33}y_3 = d_{32} = [d_2 - r_{21}d_{31}]/s_2$	
$r_{31}x_1 + r_{32}x_2 + s_3x_3 = \dots = d_{23} = d_3 - r_{31}[d_{31} - x_1] - r_{32}[d_{32} - x_2]$	
$x_3 = c_{33}y_3 = d_{33} = [d_3 - r_{31}d_{31} - r_{32}d_{32}]/s_3$	
$x_1 = c_{41}y_1 + c_{42}y_2 + c_{43}y_3 = d_{41} = d_{31} - a_{31}d_{32} - a_{31}d_{33}$	
$x_2 = c_{42}y_2 + c_{43}y_3 = d_{42} = d_{32} - a_{32}d_{33}$	
$x_3 = c_{43}y_3 = d_{43} = d_{33}$	

TABLE II

Layout of Doolittle solution for the symmetric case Coefficients not used further are indicated by ... Coefficients which can be filled by symmetry are indicated by —

$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = c_{11}y_1 + c_{12}y_2 + c_{13}y_3 = d_1$	
$— + a_{22}x_2 + a_{23}x_3 = c_{21}y_1 + c_{22}y_2 + c_{23}y_3 = d_2$	
$— — + a_{33}x_3 = c_{31}y_1 + c_{32}y_2 + c_{33}y_3 = d_3$	
$s_1x_1 + r_{21}x_2 + r_{31}x_3 = \dots = d_{21} = d_1$	
$x_1 + a_{12}x_2 + a_{13}x_3 = c_{31}y_1 + c_{32}y_2 + c_{33}y_3 = d_{31} = [d_1]/s_1$	
$s_2x_2 + r_{32}x_3 = \dots = d_{22} = d_2 - r_{21}d_{31}$	
$x_2 + a_{22}x_2 + a_{23}x_3 = c_{32}y_2 + c_{33}y_3 = d_{32} = [d_2 - r_{21}d_{31}]/s_2$	
$s_3x_3 = \dots = d_{23} = d_3 - r_{31}d_{31} - r_{32}d_{32}$	
$x_3 = \dots + c_{33}y_3 = d_{33} = [d_3 - r_{31}d_{31} - r_{32}d_{32}]/s_3$	
$x_1 = c_{41}y_1 + c_{42}y_2 + c_{43}y_3 = d_{41} = d_{31} - a_{31}d_{32} - a_{31}d_{33}$	
$x_2 = — + c_{42}y_2 + c_{43}y_3 = d_{42} = d_{32} - a_{32}d_{33}$	
$x_3 = — — + c_{43}y_3 = d_{43} = d_{33}$	

Then

$$\begin{aligned}
 B &= A - R[A_3 - I] \\
 &= A - R[(I + S^{-1}T) - I] && \text{by (4.5),} \\
 &= A - RS^{-1}T \\
 (5.2) \quad &= R + S + T. && \text{by (3.5) and (3.6).}
 \end{aligned}$$

Therefore we see that the prediagonal coefficients of the x 's in (5.1) are the elements of R and that the diagonal coefficients are the elements of S . The rest of the coefficients in this set of equations are not needed in the calculations and have been indicated by dots in Table I.

6. The symmetric case. If the A matrix is symmetric, advantage can be taken of the fact that the B matrix is also symmetric. Therefore

$$(6.1) \quad S + T = S + R'$$

and the elements of S and R' can be written down just before the division by s_{ii} in the calculation of the A_3 matrix:

$$\begin{aligned}
 (6.2) \quad A_3 &= (I + S^{-1}T) - S^{-1}(S + T) && \text{by (4.5)} \\
 &= S^{-1}(S + R')
 \end{aligned}$$

The layout of the work is given in Table II.

If the C_4 matrix is symmetric, the prediagonal elements of C_4 can be entered by symmetry. Therefore it is not only unnecessary to calculate the prediagonal elements of C_4 , but we can also omit the prediagonal elements of C_3 . Note that in this case C_4 must be calculated from the right to the left as well as from the bottom up.

The most important case where it is known in advance that C_4 is symmetric is the determination of the inverse of a symmetric matrix. Then $C' = I$ and $C_4 = A^{-1}$. Also the postdiagonal elements of C_3 are all zero so that the only elements of C_3 which have to be calculated are the diagonal elements. These are the reciprocals of the s_{ii} 's.

A case where C_4 is symmetric though $C' \neq I$ will appear in a subsequent paper.

7. Norms. In order to state the conditions for error control in a matrix calculation, a concept of the norm or the absolute value of a matrix is necessary. In this paper the norm will be defined as the square root of the sum of the squares of the elements of the matrix. That is

$$(7.1) \quad N(F) = \sqrt{\sum_i \sum_j (f_{ij})^2}.$$

The two basic inequalities satisfied by the norm are

$$(7.2) \quad N(F + G) \leq N(F) + N(G)$$

and

$$(7.3) \quad N(FG) \leq N(F)N(G).$$

All other properties of the norm are derived from these.

For future reference we list the following norm relations:

$$(7.4) \quad \begin{aligned} N[(F)(I + G)] &= N(F + FG) \\ &\leq N(F) + N(F)N(G) \\ &\leq N(F)[1 + N(G)] \end{aligned}$$

If $N(F) < 1$ we have

$$(7.5) \quad \begin{aligned} N[(I + F)^{-1} - I] &= N[I - F + F^2 - \cdots - I] \\ &\leq N(F) + [N(F)]^2 + [N(F)]^3 + \cdots \\ &\leq \frac{N(F)}{1 - N(F)}. \end{aligned}$$

If $N(G - I) < 1$, (7.5) becomes

$$(7.6) \quad 1 + N(G^{-1} - I) \leq \frac{1}{1 - N(G - I)}$$

When $N(F - I) < 1$

$$(7.7) \quad \begin{aligned} N(F^{-1}G) &\leq N[I + (F^{-1} - I)][G] \\ &\leq [1 + N(F^{-1} - I)]N(G) \\ &\leq \frac{N(G)}{1 - N(F - I)} \end{aligned} \quad \text{by (7.6)}$$

8. Error matrix and error norm. We shall also need a formal statement as to what we mean by error and we need a measure of the errors.

By an error matrix we mean the matrix whose elements consist of the differences between the value of the matrix elements as actually calculated and the true value of the matrix elements which would have been obtained if all calculations had been made exactly without any rounding or other approximations. The fundamental relation for the error matrix, $E[f(G)]$, of a function, $f(G)$, of G is

$$(8.1) \quad E[f(G)] = f[G + E(G)] - f(G).$$

If each element of a matrix is calculated to q decimal places and the matrix has p rows and p columns, the maximum rounding error introduced in any element is $.5 \times 10^{-q}$. The norm of the error introduced by rounding is less than

$$(8.2) \quad \begin{aligned} NE_1 &= \sqrt{p^2 [.5 \times 10^{-q}]^2} \\ &= .5 \times 10^{-q} \times p. \end{aligned}$$

For triangular matrices

$$(8.3) \quad NE_2 = .5 \times 10^{-9} \sqrt{p(p+1)}/2.$$

For one column matrices

$$(8.4) \quad NE_3 = .5 \times 10^{-9} \sqrt{p}.$$

For future reference the following formulas for error norms are listed:

$$(8.5) \quad NE(F+G) \leq NE(F) + NE(G).$$

$$(8.6) \quad \begin{aligned} NE(FG) &\leq NE(F)N[G+E(G)] + NE(G)N(F) \\ &\leq NE(F)N(G) + NE(G)N(F) + NE(F)NE(G). \end{aligned}$$

$$(8.7) \quad \begin{aligned} NE[F(I+G)] &\leq NE(F)[1+N[G+E(G)]] + NE(G)N(F) \\ &\leq NE(F) + NE(F)N(G) + NE(G)N(F) \\ &\quad + NE(F)NE(G). \end{aligned}$$

If $N(F-I) + NE(F) < 1$,

$$(8.8) \quad \begin{aligned} NE(F^{-1}) &= NE[I + (F-I)]^{-1} \\ &= N \left[\frac{1}{I + [F + E(F) - I]} - \frac{1}{I + (F-I)} \right] \\ &= N \{ I + [F + E(F) - I] \}^{-1} \{ I + (F-I) \}^{-1} \{ E(F) \} \\ &\leq \frac{NE(F)}{\{ 1 - N[F + E(F) - I] \} \{ 1 - N(F-I) \}} \quad \text{by (7.7),} \\ &\leq \frac{NE(F)}{[1 - N(F-I) - NE(F)] [1 - N(F-I)]}. \end{aligned}$$

If $N(F-I) + NE(F) < 1$,

$$(8.9) \quad \begin{aligned} NE(F^{-1}G) &= N \left[\frac{G + E(G)}{I + [F + E(F) - I]} - \frac{G}{I + (F-I)} \right] \\ &= N \left[\frac{[E(G)][I + (F-I)] - [E(F)][G]}{\{ I + [F + E(F) - I] \} \{ I + (F-I) \}} \right] \\ &\leq \frac{NE(G) + \{ [NE(F)]/[1 - N(F-I)] \} \{ N(G) \}}{1 - N[F + E(F) - I]} \quad \text{by (7.7).} \end{aligned}$$

9. Certain maxima. The R , S , and T matrices have no non-zero elements in common so that

$$(9.1) \quad \begin{aligned} N(B-I) &= N(R+S-I+T) \\ &= \sqrt{[N(R)]^2 + [N(S-I)]^2 + [N(T)]^2}. \end{aligned}$$

Similarly

$$(9.2) \quad NE(B) = \sqrt{[NE(R)]^2 + [NE(S)]^2 + [NE(T)]^2}.$$

The developments of the following formulas for certain maxima are not given here since they involve only well known calculus principles. It is understood that these inequalities hold whenever the quantities involved exist. Usually the maxima are given subject to the condition $N(T) = N(R)$ as well as for the unrestricted case where $N(T)$ may be zero.

$$(9.3) \quad \max \frac{N(R)N(T)}{1 - N(S - I)} = 1 - \sqrt{1 - [N(B - I)]^2}.$$

$$(9.4) \quad \max \frac{N(R)}{1 - N(S - I)} = \frac{N(B - I)}{\{1 - [N(B - I)]^2\}^{1/2}} \quad \text{if } N(T) = 0,$$

$$(9.5) \quad = \frac{N(B - I)}{\sqrt{2} \sqrt{1 - [N(B - I)]^2}} \quad \text{if } N(T) = N(R).$$

$$(9.6) \quad \max [N(R)NE(T) + N(T)NE(R)] = \sqrt{[N(B - I)]^2 - [N(S - I)]^2} \\ \times \sqrt{[NE(B)]^2 - [NE(S)]^2}.$$

Any substitution satisfying the relation

$$(9.7) \quad \frac{NE(R)}{NE(T)} = \frac{N(T)}{N(R)}$$

will cause (9.6) to attain its maximum.

$$(9.8) \quad \max[NE(R) + kNE(S)] = \sqrt{\frac{1}{2} + k^2} NE(B) \quad \text{if } NE(R) = NE(T),$$

$$(9.8.a) \quad = \sqrt{1 + k^2} NE(B) \quad \text{if } NE(T) = 0.$$

$$(9.9) \quad \max \frac{k + N(R)}{1 - N(S - I)} = k + \frac{[N(B - I)]^2}{K}$$

where K is the root of

$$(9.10) \quad [1 + k^2]K^2 + 2[k\{N(B - I)\}^2]K + \{[N(B - I)]^4 - [N(B - I)]^2\} = 0.$$

10. Errors in the Doolittle method. In all that follows, we shall assume that $N(A - I)$ is small enough so that the "divisions" are permissible. First let us examine in the multipliers $B = R + S + T$. By (3.6)

$$(10.1) \quad B - I = (A - I) - RS^{-1}T.$$

Therefore

$$N(B - I) \leq N(A - I) + \frac{N(R)N(T)}{1 - N(S - I)} \quad \text{by (7.3) and (7.7),}$$

$$(10.2) \quad \leq N(A - I) + 1 - \sqrt{1 - [N(B - I)]^2} \quad \text{by (9.3)}$$

Remembering that A has no error and letting NE_1 be the rounding error norm introduced in writing down the elements, we have

$$(10.3) \quad NE(B) \leq NE_1 + \frac{NE(R)N(T) - NE(T)N(R) + NE(R)NE(T)}{1 - N[S + E(S) - I]} \\ + \frac{NE(S)N(T)N(R)}{\{1 - N[S + E(S) - I]\}\{1 - N(S - I)\}} \quad \text{by (8.9) and (8.6).}$$

We are interested only in the range of values where the errors are small. Therefore we shall ignore second order errors. Except for such errors,

$$(10.4) \quad NE(B) \leq NE_1 + \frac{N(T)}{1 - N(S - I)} NE(R) + \frac{N(R)}{1 - N(S - I)} NE(T) \\ + \frac{N(T)N(R)}{[1 - N(S - I)]^2} NE(S).$$

The last term will be largest when $N(T) = N(R)$. Therefore the sum of the second and third will be largest when $NE(R) = NE(T)$ by (9.7).

$$(10.5) \quad NE(B) \leq NE_1 + 2 \frac{N(R)}{1 - N(S - I)} NE(R) + \left[\frac{N(R)}{1 - N(S - I)} \right]^2 NE(S).$$

By (9.8) we now obtain

$$(10.6) \quad NE(B) \leq NE_1 + K\sqrt{2} + K^2 NE(B) \\ \leq \frac{NE_1}{1 - K\sqrt{2} + K^2}$$

where

$$K = \max \frac{N(R)}{1 - N(S - I)} = \frac{N(B - I)}{\sqrt{2} \sqrt{1 - [N(B - I)]^2}} \quad \text{by (9.5).}$$

Actually in practice we introduce the rounding error in $(S^{-1}T)$ instead of in T as assumed above. Our assumption is conservative since the division by S magnifies any error in T .

Next consider the errors in the C_3 (or D_3) matrix. From (4.2)

$$(10.7) \quad C_3 = S^{-1}(C - RC_3) \\ = S^{-1}C - S^{-1}RC_3 \\ = (I + S^{-1}R)^{-1}S^{-1}C \quad \text{by (2.5) and (2.6),} \\ = [S(I + S^{-1}R)]^{-1}C \\ (10.8) \quad = (S + R)^{-1}C.$$

Therefore

$$(10.9) \quad N(C_3) \leq \frac{N(C)}{1 - N(R + S - I)} \quad \text{by (7.7),}$$

$$(10.10) \quad \leq \frac{N(C)}{1 - N(R) - N(S - I)}.$$

From (10.7), remembering that C has no error, and letting the rounding error be NE_3 ,

$$(10.11) \quad NE(C_3) \leq NE_3 + \frac{NE(R)N(C_3) + NE(C_3)N[R + E(R)]}{1 - N[S + E(S) - I]} \\ + \frac{NE(S)[N(C) + N(R)N(C_3)]}{\{1 - N[S + E(S) - I]\}\{1 - N(S - I)\}} \quad \text{by (7.7) and (7.3),} \\ \leq \frac{\frac{NE(R) + NE(S)}{1 - N(R) - N(S - I)} N(C) + \{NE_3\}\{1 - N[S + E(S) - I]\}}{1 - N[R + E(R)] - N[S + E(S) - I]}$$

since

$$(10.12) \quad \frac{N(C) + N(R)N(C_3)}{1 - N(S - I)} \\ \leq \frac{N(C)}{1 - N(S - I)} \left[1 + \frac{N(R)}{1 - N(R) - N(S - I)} \right] \quad \text{by (10.10),} \\ \leq \frac{N(C)[1 - N(S - 1)]}{[1 - N(S - I)][1 - N(R) - N(S - I)]}$$

and since transferring the terms in $NE(C_3)$ to the left requires that we divide through by

$$(10.13) \quad 1 - \frac{N[R + E(R)]}{1 - N[S + E(S) + I]} = \frac{1 - N[R + E(R)] - N[S + E(S) - I]}{1 - N[S + E(S) - I]}.$$

Again we can ignore second order errors. Taking maxima by (9.8 a) gives

$$(10.14) \quad NE(C_3) \leq \frac{\frac{\sqrt{2} NE(B)N(C)}{1 - \sqrt{2} N(B - I)} + NE_3}{1 - \sqrt{2} N(B - I)}$$

and

$$(10.15) \quad \frac{NE(C_3)}{N(C)} \leq \frac{\sqrt{2} NE(B)}{[1 - \sqrt{2} N(B - I)]^2} + \frac{[NE_3]/[N(C)]}{1 - \sqrt{2} N(B - I)}.$$

Thus we see that the proportionate error in C_3 is made up of two parts: the first due to the rounding errors in the multipliers as given by the first term and the second due to the proportionate rounding error introduced in calculating C_3 .

Finally we have the errors in the C_4 (or D_4) matrix. Since

$$(10.16) \quad C_4 = A^{-1}C$$

we have

$$(10.17) \quad N(C_4) \leq \frac{N(C)}{1 - N(A - I)}.$$

By (4.6)

$$(10.18) \quad C_4 = C_3 - (S^{-1}T)C_4.$$

If we let NE_4 be the rounding error introduced in this step

$$(10.19) \quad NE(C_4) \leq NE(C_3) + NE_4 + \frac{NE(T)N(C_4) + NE(C_4)N[T + E(T)]}{1 - N[S + E(S) - I]} \\ + \frac{NE(S)N(T)N(C_4)}{\{1 - N[S + E(S) - I]\}[1 - N(S - I)]} \\ + \{NE(C_3) + NE_4\} \{1 - N[S + E(S) - I]\} \\ \leq \frac{NE(T) + \frac{N(T)NE(S)}{1 - N(S - I)}}{1 - N[T + E(T)] - N[S + E(S) - I]} \left[\frac{N(C)}{1 - N(A - I)} \right]$$

by (10.17) and relations similar to (10.13). We now ignore second order errors and take maxima by (9.4) and (9.8.a):

$$(10.20) \quad \frac{NE(C_4)}{N(C)} \leq \frac{NE(C_3)}{N(C)} + \frac{(NE_4)}{N(C)} + \frac{\sqrt{1 + K^2} NE(B)}{1 - \sqrt{2} N(B - I)} \\ \left[\frac{1}{[1 - N(A - I)]\sqrt{1 - [N(B - I)]^2}} + \frac{\sqrt{2}}{[1 - \sqrt{2} N(B - I)]^2} \right] NE(B) \\ \leq \frac{\frac{[NE_3]/[N(C)]}{1 - \sqrt{2} N(B - I)} + \frac{NE_4}{N(C)}}{1 - \sqrt{2} N(B - I)}$$

where

$$(10.21) \quad K = \max \frac{N(T)}{1 - N(S - I)} = \frac{N(B - I)}{\sqrt{1 - [N(B - I)]^2}} \quad \text{by (9.4),}$$

and

$$(10.22) \quad \sqrt{1 + K^2} = 1/\sqrt{1 - [N(B - I)]^2}.$$

If A is symmetric, $NE(B - I)$ remains unchanged but $NE(C_4)$ can be somewhat strengthened through the use of (9.5) and (9.8) instead of (9.4) and (9.8.a).

The result is

$$(10.23) \quad \frac{NE(C_4)}{N(C)} \leq \left[\frac{1}{[1 - N(A - I)]\sqrt{2}\sqrt{1 - [N(B - I)]^2}} + \frac{\sqrt{3/2}}{[1 - \sqrt{3/2} N(B - I)]^2} \right] \left[\frac{NE(B)}{1 - \sqrt{3/2} N(B - I)} \right] \\ + \frac{1}{[1 - \sqrt{3/2} N(B - I)]^2} \frac{NE_3}{N(C)} \\ + \frac{1}{1 - \sqrt{3/2} N(B - I)} \frac{NE_4}{N(C)}.$$

If C is approximately equal to I , a better formula is obtained by substituting $(I + C_1)$ for C . The final formulas then are identical to (10.20) and (10.23) with the substitution of $1 + N(C_1)$ for $N(C)$. Similarly if $C = I$ so that $C_4 = A^{-1}$, $N(C) = 1$ should be substituted in (10.20) and (10.23).

If A is symmetric and if $C = I$ so that $C_4 = A^{-1}$, the prediagonal elements of C_4 are filled in by symmetry as in Table II instead of being calculated directly. This complicates the analysis of error relations. The following inequality gives the error limit for the diagonal and postdiagonal elements of (A^{-1}) . We have indicated these elements by F .

$$(10.24) \quad NE(F) \leq \frac{\sqrt{[3/2] - N(A - I)} + [N(B - I)]^2/K}{[1 - N(A - I)][1 - \sqrt{2} N(B - I)]} NE(B) \\ + \frac{NE_5}{1 - \sqrt{2} N(B - I)}$$

where K is the root of (9.10) when $k = 1 - N(A - I)$ and $NE_5 = 0.5 \times 10^{-q} \times \sqrt{p(p+1)/2}$ by (8.3)

11. Results. Given the matrix A , we subtract one from each element on the principal diagonal to obtain the matrix $(A - I)$. By the norm of $(A - I)$ we mean the square root of the sum of the squares of the elements of $(A - I)$. We shall now show that a Doolittle process such as outlined in Table I is in a state of error control if the norm of $(A - I)$ is less than 0.35:

1. $N(A - I) \leq 0.35$
2. $N(B - I) \leq 0.4642$ by (10.2).
3. $NE(B) \leq 1.09 p$ by (10.6) and (8.2)

if the maximum rounding error in any element is 0.5 and A has p rows. Thus no element of the multiplying matrices R , S , or T can have an error of greater than this amount.

4. $NE(C_4) \leq (44 \times p \times 10^{-q})N(C) + (6 \times p \times 10^{-r})$ by (10.20) and (8.2) where q decimal places are carried on the left and r on the right. Thus if the decimal point in C is shifted so that $N(C) \leq 1$, the error in any element of C_4 can

not amount to more than three significant figures if the same number of decimal places are carried on both the left and on the right (four significant figures for 21 to 200 rows).

5. $N(A^{-1})$ \leq three significant figures by substituting $N(C) = 1$ since $C = I$.

As we let $N(A + I)$ become larger than 0.35, the maximum errors indicated by our formulas rapidly become very large. In fact they become infinite if $N(A + I) = 0.111$.

Since for more than four equations the above formulas show errors in the second decimal place no matter how small $N(A + I)$ is, it is suggested that as a general practice:

1. The problem be arranged so that $N(A + I) \leq 0.35$.
2. The decimal points in C be shifted so $N(C) = 1$.
3. Three extra decimal places be carried in the calculations.

12. Preliminary adjustments. The requirement that $N(A + I)$ should be less than 0.35 is not normally met in practical problems. If, however, an approximation to A^{-1} is available, the problem can almost always be rearranged to satisfy this condition.

Thus if we are solving the equations such as given in Table I,

$$(12.1) \quad AX = CY = D,$$

we are perfectly free to multiply through by any non-singular matrix F without disturbing the solution:

$$(12.2) \quad (FA)X = (FC)Y = (FD)$$

Now if F is a sufficiently close approximation to A^{-1} , FA will be almost equal to I . Therefore $N(FA + I)$ will be less than 0.35 and a Doolittle solution of (12.2) will be in a state of error control.

Similarly for the inversion of A , we can apply the Doolittle process to the pair of matrices

$$FA : F$$

just as easily as to the pair

$$A : I$$

since

$$\begin{aligned} (FA)^{-1} \|FA : F\| &= A^{-1} F^{-1} \|FA : F\| \\ &= \|A^{-1} (F^{-1} F) A : A^{-1} (F^{-1} F)\| \\ &= \|A^{-1} A : A^{-1}\| \\ &= \|I : A^{-1}\|. \end{aligned}$$

Thus by taking F as a sufficiently close approximation to A^{-1} , we can bring an inversion calculation into a state of error control.

The computer should be cautioned that the multiplication by F must be exact and that no rounding is allowable in this step. Our formulas assumed that we started with matrices free of error.

13. Further work. The principles used in this paper can be applied to the task of developing calculation routines which will be in a state of error control regardless of the size of $N(A - I)$. Enough work has been done to see that such routines do exist and do not involve prohibitive labor. The author expects that the most efficient routine will be to use these more elaborate methods to obtain an F such that $N(FA - I) \leq 0.35$ and then to use the normal Doolittle methods as outlined in section 12.

REFERENCES

- The writer wishes to acknowledge particularly his debt to Prof. Hotelling's very complete paper on interactive methods and to Prof. Dwyer's papers on the Doolittle methods. These papers have such complete bibliographies, we shall not give further references here.
- [1] PAUL S. DWYER, "The Doolittle technique," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 449-458.
 - [2] PAUL S. DWYER, "The solution of simultaneous equations," *Psychometrika*, Vol. 6 (1941), pp. 101-129.
 - [3] HAROLD HOTELLING, "Some new methods in matrix calculation," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 1-34.

INVERSE TABLES OF PROBABILITIES OF ERRORS OF THE SECOND KIND

BY EMMA LEHMER

University of California, Berkeley

1. Introduction. The problem of testing linear hypotheses was discussed by Kolodziejczyk [1], and later in greater detail by Tang [2], who computed a table giving the probabilities of errors of the second kind P_{II} for a range of values of the two degrees of freedom f_1 and f_2 , ($f_1 = 1(1)8, f_2 = 2, 4, 6(1)30, 60, \infty$)¹ and for two fixed levels $P_I = .01$ and $.05$ of the probability of errors of the first kind. These tables are in terms of a parameter φ , ($\varphi = 1(.5)3(1)8$) whose statistical significance, or rather that of

$$\lambda = (f_1 + 1)\varphi^2/2$$

is discussed in Tang's paper. A restatement of the problem of testing linear hypotheses in a more canonical form, giving an interpretation of λ , will also be found in a recent paper by Wald [3].

Professor Neyman has felt for some time that a table giving $\varphi = \varphi(a, b, \alpha, \beta)$ as a function of the two degrees of freedom $f_1 = 2a$, and $f_2 = 2b$, and of the two probability levels $\alpha = P_I$ and $\beta = 1 - P_{II}$ would be more useful for statistical purposes, where β is the probability of detecting the falsehood of the hypothesis tested. A paper by Professor Neyman explaining this point of view and giving applications of the present tables to some statistical problems will appear shortly. These tables were computed in the Statistical Laboratory of the University of California,² and give values of φ for the following range of parameters:

$$(\alpha, \beta) = (.01, .7), (.01, .8), (.05, .7), (.05, .8)$$

$$f_1 = 1(1)10, 12, 15, 20, 24, 30, 40, 60, 80, 120, \infty.$$

$$f_2 = 2(2)20, 24, 30, 40, 60, 80, 120, 240, \infty.$$

2. Analytic definitions. The statistical parameter

$$(1) \quad \lambda = \lambda(a, b, \alpha, \beta) = (a + \frac{1}{2}) \varphi^2(a, b, \alpha, \beta)$$

can be thought of as an inverse function connected with the hypergeometric distribution. Inverse functions $y(\alpha)$, $u(a, \alpha)$ and $x(a, b, \alpha)$ of the better known normal, Gamma and Beta distributions respectively have all been tabulated,

* ¹ The notation $m = r(s)l$ is equivalent to $m = r, r + s, r + 2s, \dots, l$.

² These tables were begun by Miss Leone Gintzler, and were carried on by Mark Eudy under a University of California Research Grant. The bulk of the computing was done, however, by the author and by Mrs. Julia Robinson under a grant of the American Philosophical Society.

and are sometimes called "percentage points" of the distribution. To begin with the simplest, the normal distribution, we may define $y(\alpha)$ as the solution of

$$(2) \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y(\alpha)} e^{-x^2/2} dx = \alpha.$$

This function has been recently tabulated by Truman Lee Kelley [4] for $\alpha = p = 0.0001$ 1 to 8 decimal places.

The function $u(a, \alpha)$ is the solution of

$$(3) \quad \frac{1}{\Gamma(a)} \int_0^{u(a, \alpha)} t^{a-1} e^{-t} dt = \Gamma_u(a) / \Gamma(a) = \alpha.$$

This is connected with the well known percentage points of the χ^2_ν distribution with $\nu = 2a$ degrees of freedom as follows:

$$(4) \quad \chi^2_\nu(\alpha) = 2u(a, 1 - \alpha).$$

Catherine Thomson [5] has tabulated $\chi^2_\nu(\alpha)$ for $\alpha = .005, .01, .025, .05, .1, .25, .5, .75, .9, .95, .975, .99, .995$, and for $\nu = 1(1)30(10)100$. She has also tabulated [6] the corresponding parameter $x = x(a, b, \alpha)$ of the Beta distribution with $\nu_1 = 2b, \nu_2 = 2a$ degrees of freedom defined by

$$(5) \quad \frac{1}{B(a, b)} \int_0^{x(a, b, \alpha)} t^{a-1} (1-t)^{b-1} dt = \frac{B_x(a, b)}{B(a, b)} = \alpha$$

for $\alpha = .005, .01, .025, .05, .1, .25$ and $.5$ and $\nu_1 = 1(1)10, 12, 15, 20, 24, 30, 40, 60, 120, \nu_2 = 1(1)30, 40, 60, 120, \infty$ to five significant places.

Similarly $\lambda(a, b, \alpha, \beta)$ can be defined as the solution of

$$(6) \quad \frac{1}{B(a, b)} \int_0^{x(a, b, 1-\alpha)} e^{-\lambda(1-t)} t^{a-1} (1-t)^{b-1} F(-b, a, -\lambda t) dt = 1 - \beta$$

where

$$F(\gamma, \delta, z) = 1 + \frac{\gamma}{\delta} z + \frac{\gamma(\gamma+1)}{\delta(\delta+1)} z^2 + \dots$$

is the confluent hypergeometric function.

3. Limiting cases. It is well known that as a tends to infinity

$$(7) \quad \frac{\chi^2_\nu(\alpha)}{\nu} = \frac{u(a, 1 - \alpha)}{a} = 1 + \frac{y(\alpha)}{\sqrt{a}} + o\left(\frac{1}{a}\right).$$

There are many approximations [7] to χ^2 . In a recent paper Peiser [8] gave a rigorous derivation of an asymptotic formula for χ^2 .

Similarly, the limiting cases of $x(a, b, \alpha)$ as a and b tend to infinity are known

TABLE OF φ Level of Significance, $\alpha = .05$. Probability of Detecting the Falsehood of the Hypothesis Tested, $\beta = .7$

$f_2 \backslash f_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60	120	∞	240 ^{1/2}
2	3.438	3.920	4.141	4.267	4.350	4.408	4.451	4.484	4.510	4.532	4.564	4.598	4.632	4.650	4.667	4.685	4.704	4.722	4.740	
4	2.341	2.418	2.436	2.441	2.442	2.442	2.441	2.440	2.439	2.438	2.437	2.434	2.432	2.43	2.429	2.43	2.426	2.424	2.421	
6	2.104	2.091	2.059	2.031	2.010	1.992	1.978	1.966	1.956	1.948	1.935	1.921	1.905	1.90	1.888	1.88	1.870	1.860	1.850	
8	2.003	1.950	1.895	1.852	1.818	1.792	1.771	1.754	1.739	1.727	1.707	1.685	1.662	1.65	1.636	1.62	1.608	1.593	1.576	
10	1.948	1.872	1.803	1.751	1.710	1.679	1.653	1.632	1.614	1.599	1.575	1.548	1.519	1.50	1.487	1.47	1.451	1.431	1.411	
12	1.913	1.823	1.745	1.687	1.641	1.606	1.577	1.553	1.533	1.516	1.488	1.457	1.424	1.41	1.388	1.37	1.341	1.321	1.297	
14	1.888	1.789	1.705	1.642	1.593	1.554	1.523	1.497	1.475	1.457	1.426	1.393	1.356	1.34	1.314	1.29	1.267	1.240	1.212	
16	1.871	1.764	1.675	1.609	1.557	1.517	1.483	1.456	1.432	1.413	1.380	1.344	1.304	1.28	1.259	1.23	1.206	1.178	1.146	
18	1.857	1.744	1.653	1.584	1.530	1.487	1.453	1.424	1.400	1.378	1.344	1.306	1.264	1.24	1.215	1.19	1.158	1.127	1.093	
20	1.846	1.729	1.635	1.563	1.508	1.464	1.428	1.398	1.373	1.351	1.316	1.276	1.231	1.21	1.180	1.15	1.120	1.086	1.048	12
24	1.831	1.707	1.608	1.534	1.476	1.430	1.392	1.360	1.333	1.310	1.273	1.230	1.182	1.16	1.13	1.09	1.06	1.02	978	10
30	1.815	1.685	1.582	1.505	1.444	1.396	1.356	1.322	1.294	1.269	1.229	1.183	1.131	1.10	1.070	1.03	996	951	902	8
40	1.80	1.66	1.56	1.48	1.41	1.36	1.32	1.28	1.25	1.23	1.19	1.14	1.08	1.05	1.01	.97	93	88	816	6
60	1.786	1.643	1.532	1.447	1.381	1.328	1.284	1.247	1.215	1.187	1.141	1.088	1.026	99	951	90	854	792	714	4
80	1.78	1.63	1.52	1.43	1.37	1.31	1.27	1.23	1.20	1.17	1.12	1.06	1.00	.96	92	.87	81	75	653	3
120	1.77	1.62	1.51	1.42	1.35	1.29	1.25	1.21	1.18	1.15	1.10	1.04	.97	93	87	.83	.77	692	577	2
240	1.76	1.61	1.50	1.41	1.33	1.28	1.23	1.19	1.16	1.13	1.07	1.01	.94	.90	.85	.79	.72	63	473	1
∞	1.757	1.602	1.483	1.392	1.320	1.262	1.213	1.170	1.135	1.104	1.051	.988	.913	.87	.815	.75	.673	557	0	0

TABLE OF φ Level of Significance, $\alpha = 0.5$. Probability of Detecting the Falsehood of the Hypothesis Tested, $\beta = 8$

$f_2 \backslash f_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	$240/f_2$
2	3.598	4.558	4.814	4.961	5.057	5.125	5.175	5.213	5.244	5.269	5.307	5.346	5.386	5.406	5.427	5.447	5.469	5.490	5.511	
4	2.650	2.736	2.752	2.755	2.755	2.754	2.752	2.750	2.749	2.747	2.745	2.742	2.739	2.74	2.735	2.73	2.730	2.728	2.725	
6	2.381	2.351	2.309	2.274	2.248	2.226	2.200	2.165	2.183	2.174	2.153	2.141	2.122	2.11	2.103	2.09	2.081	2.069	2.053	
8	2.263	2.188	2.119	2.066	2.026	1.995	1.970	1.949	1.932	1.918	1.891	1.869	1.841	1.83	1.811	1.80	1.779	1.761	1.742	
10	2.199	2.098	2.014	1.951	1.903	1.865	1.835	1.810	1.789	1.771	1.743	1.712	1.678	1.66	1.640	1.62	1.599	1.577	1.553	
12	2.159	2.041	1.947	1.877	1.824	1.782	1.748	1.720	1.696	1.676	1.644	1.609	1.570	1.55	1.526	1.50	1.478	1.452	1.423	
14	2.131	2.002	1.901	1.826	1.769	1.723	1.687	1.656	1.631	1.609	1.574	1.535	1.493	1.47	1.444	1.42	1.390	1.360	1.328	
16	2.110	1.974	1.868	1.789	1.730	1.681	1.642	1.610	1.583	1.560	1.522	1.481	1.434	1.41	1.382	1.35	1.322	1.289	1.253	
18	2.095	1.952	1.842	1.760	1.698	1.647	1.607	1.574	1.545	1.521	1.482	1.438	1.389	1.36	1.333	1.30	1.268	1.233	1.193	
20	2.083	1.935	1.822	1.737	1.673	1.621	1.580	1.545	1.516	1.490	1.449	1.404	1.352	1.32	1.294	1.26	1.225	1.186	1.143	12
24	2.065	1.910	1.792	1.704	1.636	1.583	1.539	1.501	1.471	1.444	1.401	1.352	1.297	1.27	1.23	1.20	1.16	1.11	1.065	10
30	2.047	1.886	1.763	1.671	1.601	1.544	1.498	1.460	1.427	1.399	1.352	1.300	1.240	1.21	1.171	1.13	1.087	1.037	979	8
40	2.03	1.86	1.73	1.64	1.57	1.51	1.46	1.42	1.38	1.35	1.30	1.25	1.18	1.15	1.11	1.06	1.01	95	885	6
60	2.014	1.838	1.706	1.607	1.531	1.469	1.418	1.375	1.339	1.307	1.254	1.194	1.124	1.08	1.039	99	930	857	773	4
80	2.01	1.83	1.69	1.59	1.51	1.45	1.40	1.35	1.32	1.28	1.23	1.17	1.09	1.05	1.00	95	88	81	705	3
120	2.00	1.81	1.68	1.58	1.50	1.43	1.38	1.33	1.30	1.26	1.21	1.14	1.06	1.02	97	91	84	750	623	2
240	1.99	1.80	1.67	1.56	1.48	1.41	1.36	1.31	1.27	1.24	1.18	1.11	1.03	98	93	87	79	68	509	1
∞	1.981	1.792	1.654	1.545	1.462	1.395	1.339	1.292	1.251	1.215	1.155	1.084	997	95	889	82	731	603	0	0

TABLE OF ϕ Level of Significance, $\alpha = .01$. Probability of Detecting the Hypothesis Tested, $\beta = .8$

$f_1 \backslash f_2$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
2	8.905	10.326	10.943	11.297	11.527	11.689	11.809	11.901	11.974	12.034	12.125	12.219	12.314	12.363	12.412	12.462	12.512	12.563	12.613
4	4.157	4.311	4.353	4.372	4.380	4.383	4.385	4.386	4.386	4.386	4.386	4.386	4.384	4.383	4.383	4.381	4.380	4.379	4.378
6	3.371	3.335	3.282	3.239	3.205	3.179	3.157	3.140	3.125	3.113	3.093	3.072	3.050	3.037	3.023	3.012	2.998	2.984	2.969
8	3.068	2.959	2.866	2.793	2.746	2.706	2.674	2.647	2.625	2.607	2.577	2.545	2.511	2.492	2.473	2.453	2.432	2.410	2.388
10	2.910	2.763	2.649	2.566	2.503	2.455	2.416	2.384	2.358	2.335	2.299	2.260	2.218	2.195	2.171	2.146	2.120	2.092	2.063
12	2.813	2.642	2.515	2.423	2.353	2.300	2.256	2.220	2.191	2.166	2.125	2.081	2.032	2.006	1.978	1.950	1.919	1.886	1.851
14	2.748	2.561	2.425	2.327	2.252	2.194	2.148	2.109	2.077	2.049	2.005	1.957	1.904	1.875	1.844	1.812	1.778	1.741	1.701
16	2.701	2.503	2.358	2.257	2.179	2.118	2.069	2.028	1.994	1.965	1.918	1.866	1.809	1.778	1.745	1.710	1.672	1.632	1.588
18	2.666	2.460	2.312	2.205	2.124	2.060	2.009	1.966	1.931	1.901	1.851	1.796	1.736	1.703	1.668	1.630	1.590	1.546	1.498
20	2.638	2.426	2.274	2.164	2.080	2.014	1.962	1.918	1.881	1.850	1.798	1.741	1.679	1.644	1.607	1.567	1.524	1.477	1.425
24	2.60	2.38	2.22	2.10	2.02	1.95	1.89	1.85	1.81	1.77	1.72	1.66	1.59	1.56	1.52	1.47	1.42	1.37	1.312
30	2.559	2.328	2.164	2.046	1.955	1.884	1.826	1.778	1.737	1.702	1.644	1.580	1.507	1.47	1.421	1.37	1.321	1.261	1.193
40	2.52	2.28	2.11	1.99	1.89	1.82	1.76	1.71	1.67	1.63	1.57	1.50	1.42	1.38	1.33	1.27	1.21	1.14	1.064
60	2.485	2.236	2.062	1.935	1.837	1.760	1.696	1.643	1.598	1.559	1.494	1.420	1.335	1.29	1.232	1.17	1.102	1.019	.917
80	2.47	2.21	2.04	1.91	1.81	1.73	1.66	1.61	1.56	1.52	1.46	1.38	1.29	1.24	1.18	1.12	1.04	.95	.830
120	2.45	2.19	2.01	1.88	1.78	1.70	1.63	1.58	1.53	1.49	1.42	1.34	1.25	1.20	1.13	1.06	.98	.875	.727
240	2.43	2.17	1.99	1.86	1.75	1.67	1.60	1.55	1.50	1.46	1.38	1.30	1.20	1.15	1.08	1.00	.91	.79	.587
∞	2.417	2.150	1.966	1.830	1.726	1.642	1.572	1.514	1.464	1.422	1.346	1.256	1.158	1.10	1.026	.94	.839	.687	0

TABLE OF φ Level of Significance, $\alpha = .01$. Probability of Detecting the Falsehood of the Hypothesis Tested, $\beta = 7$

$f_1 \backslash f_2$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	$240/f_2$
2	7.746	8.922	9.455	9.761	9.960	10.094	10.204	10.282	10.346	10.398	10.476	10.557	10.640	10.681	10.724	10.767	10.810	10.855	10.900	12
4	3.714	3.858	3.900	3.917	3.923	3.929	3.931	3.932	3.933	3.933	3.934	3.933	3.932	3.932	3.931	3.930	3.929	3.928	3.928	10
6	3.037	3.015	2.972	2.935	2.907	2.884	2.865	2.850	2.838	2.827	2.810	2.790	2.771	2.761	2.750	2.738	2.726	2.713	2.700	8
8	2.773	2.686	2.608	2.548	2.503	2.468	2.440	2.416	2.398	2.381	2.356	2.327	2.297	2.281	2.264	2.246	2.227	2.208	2.188	6
10	2.634	2.512	2.415	2.343	2.289	2.246	2.212	2.184	2.160	2.141	2.109	2.074	2.037	2.016	1.995	1.973	1.950	1.925	1.899	4
12	2.548	2.406	2.296	2.216	2.155	2.108	2.070	2.038	2.012	1.990	1.954	1.914	1.871	1.848	1.824	1.798	1.770	1.740	1.710	3
14	2.490	2.334	2.216	2.130	2.064	2.013	1.972	1.938	1.910	1.886	1.847	1.804	1.756	1.730	1.703	1.674	1.645	1.611	1.575	2
16	2.448	2.282	2.158	2.068	1.999	1.945	1.901	1.865	1.834	1.807	1.763	1.722	1.671	1.643	1.614	1.582	1.548	1.512	1.473	1
18	2.417	2.243	2.114	2.020	1.949	1.893	1.847	1.810	1.778	1.752	1.708	1.659	1.605	1.577	1.545	1.511	1.474	1.435	1.392	0
20	2.392	2.212	2.080	1.984	1.910	1.852	1.805	1.767	1.734	1.706	1.660	1.610	1.553	1.525	1.489	1.453	1.414	1.372	1.326	
24	2.336	2.167	2.030	1.930	1.853	1.79	1.74	1.70	1.67	1.64	1.59	1.54	1.48	1.44	1.40	1.36	1.32	1.28	1.223	
30	2.321	2.124	1.982	1.877	1.797	1.734	1.682	1.639	1.603	1.571	1.520	1.462	1.397	1.36	1.321	1.28	1.230	1.176	1.115	
40	2.29	2.08	1.94	1.83	1.74	1.68	1.62	1.58	1.54	1.50	1.45	1.39	1.32	1.28	1.24	1.19	1.13	1.07	.996	
60	2.255	2.041	1.889	1.776	1.690	1.621	1.564	1.517	1.476	1.441	1.383	1.316	1.239	1.20	1.147	1.09	1.029	.954	.860	
80	2.24	2.02	1.87	1.75	1.66	1.59	1.54	1.49	1.44	1.41	1.35	1.28	1.20	1.16	1.10	1.04	.97	.89	.780	
120	2.22	2.00	1.84	1.73	1.64	1.57	1.51	1.46	1.41	1.38	1.32	1.24	1.16	1.11	1.05	.99	.91	.820	.684	
240	2.21	1.98	1.82	1.70	1.61	1.54	1.48	1.43	1.38	1.35	1.28	1.21	1.12	1.07	1.01	.94	.85	.74	.560	
∞	2.193	1.963	1.801	1.678	1.587	1.513	1.451	1.400	1.353	1.313	1.247	1.169	1.076	1.02	.956	.882	.785	.647	0	

[9] and follow readily from (5), although no attempt has been made, as far as I know, to find better approximations. These limiting values are as follows:

$$(8) \quad \lim_{a \rightarrow \infty} a[1 - x(a, b, \alpha)] = u(b, 1 - \alpha)$$

$$(9) \quad \lim_{b \rightarrow \infty} bx(a, b, \alpha) = u(a, \alpha).$$

The two corresponding limiting cases for λ are not at first glance so symmetric. When a tends to infinity, we have from (1)

$$(10) \quad \lim_{a \rightarrow \infty} \frac{\lambda}{a} = \varphi^2$$

while

$$(11) \quad \lim_{\alpha \rightarrow \infty} F(-b, a, -a\varphi^2 t) = (1 + \varphi^2 t)^b$$

Substituting these in (6) and letting $t = 1 - z/a(1 + \varphi^2)$ and passing to the limit we get with the help of (8)

$$(12) \quad \frac{1}{\Gamma(b)} \int_{(1+\varphi^2)u(b,\alpha)}^{\infty} c^{-1} z^{b-1} dz = 1 - \beta.$$

In other words

$$(1 + \varphi^2)u(b, \alpha) = u(b, \beta)$$

or

$$(13) \quad \varphi(\infty, b, \alpha, \beta) = \sqrt[b]{\frac{u(b, \beta)}{u(b, \alpha)}} - 1.$$

This is the only case, except for $b = 1$ in which φ can be given explicitly. For $b = 1$, we have from (5)

$$x(a, 1, \alpha) = \sqrt[a]{\alpha}$$

and (6) can be easily integrated to give

$$\varphi(a, 1, \alpha, \beta) = \left[\log \left(\frac{1 - \alpha}{1 - \beta} \right) / \left(a + \frac{1}{2} \right) (1 - \sqrt[a]{1 - \alpha}) \right]^{\frac{1}{a-1}}.$$

In all other cases it was found impractical to attempt an inversion of (6) to get φ .

When b becomes infinite (6) becomes with $t = z/b$, and with the help of (9),

$$(14) \quad e^{-\lambda} \int_0^{u(a, 1-\alpha)} e^{-z} \left(\frac{z}{\lambda} \right)^{(a-1)/2} I_{a-1}(2\sqrt{\lambda z}) dz = 1 - \beta$$

where

$$(15) \quad I_{a-1}(2\sqrt{\lambda z}) = \frac{(\lambda z)^{-(a-1)/2}}{\Gamma(a)} \lim_{b \rightarrow \infty} F\left(-b, a, -\frac{\lambda z}{b}\right)$$

is the Bessel function of a purely imaginary argument, which is usually defined by

$$I_n(x) = \sum_{\nu=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{n+2\nu}}{\nu! \Gamma(n + \nu + 1)}.$$

Expression (14) was also obtained for this limiting case of the hypergeometric distribution by Wishart [10]. This integral, however, does not give λ any more explicitly than the general integral (6), and since the calculation of λ increases in difficulty as a increases, an attempt was made to derive an approximate formula for $\varphi(a, \infty, \alpha, \beta)$ for large a . To this end an asymptotic formula [11] was developed for $I_n(nx)$, the principal term of which is

$$(16) \quad I_n(nx) = \frac{(1+x^2)^{1/4}}{\sqrt{2\pi n}} \left(\frac{\sqrt{1+x^2}-1}{x} e^{\sqrt{1+x^2}} \right)^n.$$

Substituting this into (14) with $x = 2\sqrt{\lambda z/a} - 1$ and $n = a - 1$ we get for large a

$$(17) \quad \frac{e^{-\lambda}}{\Gamma(a)} \int_0^{u(a, 1-\alpha)} t^{a-1} e^{-t(1-\varphi^2+t\varphi^4/2a)} dt = 1 - \beta$$

If we assume that φ is sufficiently small to neglect the term in φ^4 we get as a first approximation

$$(18) \quad \frac{1}{\Gamma(a)} \int_0^{(1-\varphi^2)u(a, 1-\alpha)} t^{a-1} e^{-t} dt = 1 - \beta$$

or

$$(19) \quad \varphi(a, \infty, \alpha, \beta) \sim \sqrt{1 - \frac{u(a, 1-\beta)}{u(a, 1-\alpha)}}$$

a formula very similar to (11). In fact since a is large this formula can be reduced one more step using (7). This gives

$$(20) \quad \lim_{a \rightarrow \infty} \sqrt{a} \varphi^2(a, \infty, \alpha, \beta) = y(\alpha) - y(\beta)$$

Similarly (13) becomes

$$\lim_{b \rightarrow \infty} \sqrt{a} \varphi^2(\infty, b, \alpha, \beta) = y(\alpha) - y(\beta).$$

If instead of neglecting the term in φ^4 , we multiply it by the value of t at its upper limit, we get

$$(21) \quad \varphi^2 \sim \frac{1 - \sqrt{1 - 2[u(a, 1-\alpha) - u(a, 1-\beta)]/a}}{u(a, 1-\alpha)/a}.$$

Professor Neyman derived another approximation for $\varphi(a, \infty, \alpha, \beta)$ by assuming that the distribution (14) approaches a normal distribution for large a . He obtained:

$$(22) \quad \sqrt{a} \varphi^2 \sim y(\alpha) + \frac{y^2(\beta)}{\sqrt{a}} - y(\beta) \left(1 + \frac{2y(\alpha)}{\sqrt{a}} + \frac{y^2(\beta)}{a} \right)^{1/2}.$$

Both (21) and (22) obviously reduce to (20) in the limit. The following table shows the efficiency of these formulas for $a = 60$

	Table	(21)	(22)	(20)
$\begin{cases} \alpha = .01 \\ \beta = .8 \end{cases}$.687	.695	.668	.640
$\begin{cases} \alpha = .01 \\ \beta = .7 \end{cases}$.647	.642	.622	.607
$\begin{cases} \alpha = .05 \\ \beta = .8 \end{cases}$.603	.585	.593	.566
$\begin{cases} \alpha = .05 \\ \beta = .7 \end{cases}$.557	.540	.544	.529

A rigorous derivation of some such formula giving the actual order of approximation of φ would of course be of interest, but is likely to be quite complicated.

4. Calculation of tables. It is fairly obvious that the integral (6), although very useful theoretically is not well adapted to actual calculations. It can easily be integrated by parts to produce the infinite series.

$$(23) \quad e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \frac{B_{1-x}(a+i, b)}{B(a+i, b)} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \frac{B_x(b, a+i)}{B(a+i, b)} = 1 - \beta.$$

This series can be used effectively for calculation purposes only if λ is comparatively small. If b is an integer, however, this series can be replaced by a finite series of b terms, which was also used by Tang [2] in calculating his tables. This series is as follows with $x = x(a, b, 1 - \alpha) = x(b, a, \alpha)$:

$$(24) \quad e^{-\lambda x} (1-x)^{a+b-1} \sum_{i=0}^{b-1} T_i = 1 - \beta$$

where

$$T_0 = 1, \quad T_1 = x[\lambda(1-x) + a + b - 1]/(1-x)$$

and³

$$(25) \quad T_n = x\{[\lambda(1-x) + a + b - n] T_{n-1} + \lambda x T_{n-2}\}/n(1-x).$$

The subjoined tables can be thought of as inverses of Tang's tables, and could have been obtained from tables such as Tang's by inverse interpolation, had the interval of tabulation been sufficiently fine. The interval of tabulation of .5 for φ allowed only a crude approximation or trial value of φ , the corresponding probability was then calculated for this point, and then corrected with the help

³ It will be noticed in comparing these formulas with those given by Tang, that x is used for $1 - x$. This is done to conform with Miss Thomson's table for x .

of derivatives. In the beginning of the work, a recalculation was usually made for the corrected value of φ , and the tabulated value of φ was then obtained by inverse interpolation between these very close values. As the work progressed, and difference tables were calculated in several directions, the guesses improved considerably so that a correction could be made using the first derivative to give a tabulated value of φ correct to three decimal places. Such corrections never exceeded .004, and therefore it is hoped that the tables are correct to the last place. The derivative in question is given by

$$(26) \quad \frac{d\beta}{d\varphi} = (2a + 1)\varphi x(1 - x)^{a+b-1}e^{-\lambda x}T_{b-1},$$

and was obtained as a by-product of the calculation of (24). This method was used for all finite a 's and for all b 's less than 30. For $b = 30$ or more it was found more expeditious to use the infinite series (23), about 20 terms of which sufficed.

The values of $x(a, b, 1 - \alpha) = x(b, a, \alpha)$ used in these calculations were obtained to five significant places from Miss Thomson's table [6] for $\alpha = .01$ and .05 with $\nu_1 = 2a$, $\nu_2 = 2b$.

No calculations were made for non-integer b 's since for small values of b , λ was too large to make the infinite series (23) practicable, while for $b \geq 7$, φ can be easily obtained by interpolation. The only available method for calculating φ for $b = 1/2, 3/2$ or $5/2$ would be by numerical integration of (6), which would be rather lengthy. Furthermore, the interest seems to be in large rather than small values of a and b .

5. Calculations for infinite cases. The case $a = \infty$ was readily disposed of using (13), as for $b = \infty$, the integral (14) was again integrated by parts to give

$$(27) \quad e^{-(\lambda+u)} \sum_{\nu=0}^{\infty} \left(\frac{u}{\lambda}\right)^{(a+\nu)/2} I_{a+\nu}(2\sqrt{u\lambda}) = 1 - \beta.$$

This was found to be effective, especially when $u < \lambda$, which is the case for small values of a . When u exceeded λ , the complementary series was used, namely

$$(28) \quad e^{-(\lambda+u)} \sum_{\nu=0}^{\infty} \left(\frac{u}{\lambda}\right)^{(a-1-\nu)/2} I_{a-1-\nu}(2\sqrt{u\lambda}) = \beta.$$

The calculations proceeded in much the same manner as in the finite case. The values of $2u$ were obtained from Miss Thomson's table of percentage points of $\chi^2_\nu(\alpha)$ distribution with $\nu = 2a$ degrees of freedom, while the values of I_0 and I_1 were obtained from the tables of Bessel Functions [12] for $2\sqrt{\lambda z} \leq 20$, and from the tables of Anding [13] for larger values of the variable. The values of I_ν for $\nu > 1$, were computed from the recurrence relation⁴

⁴ If a is an odd integer, the values of $I_\nu(z)$ can be built up using (29) from $I_{-1}(z) = \frac{\cosh z}{\sqrt{\pi z}}$

and $I_{\frac{1}{2}}(z) = \frac{\sinh z}{\sqrt{\pi z}}$, which of course are tabled.

$$(29) \quad I_{r+1} = -\frac{2\nu}{z} I_r(z) + I_{r-1}(z).$$

Future calculations of this sort could make use of the forthcoming tables [14] of $I_\nu(z)$ for $\nu \leq 20$, and $z \leq 20$ to ten significant places. As before, the tabulated values were obtained by correcting trial values of φ by means of the derivative

$$(30) \quad \frac{d\beta}{d\varphi} = \sqrt{2u(2a+1)} e^{-(\lambda+u)} \left(\frac{u}{\lambda}\right)^{(a-1)/2} I_a(2\sqrt{u\lambda}),$$

which again was obtained as a by-product of the calculations. This method becomes impractical when a is too large, because a great deal of accuracy is lost in applying recurrence (29) many times. For some of the larger values of a it was found preferable to use the series

$$(31) \quad 1 - \beta = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \frac{\Gamma_u(a+i)}{\Gamma(a+i)}$$

although it converges rather slowly. In other words the upper limits for a and b were pushed as far as was practicable.

The values which are tabled to two instead of three decimals were interpolated using second differences, all other values were computed in the manner described above. Difference tables were made by rows, columns and between tables, as a final check on the work. Difference tables using harmonic interpolation were also made for both rows and columns, and found very effective, with the exception of the lower-right hand corner, where φ drops rapidly to zero. The last column of each table is to be used for harmonic interpolation.

REFERENCES

- [1] S. KOŁODZIEJCZYK, *Biometrika*, V. 27 (1935), p. 161
- [2] P. C. TANG, *Stat. Res. Memoirs*, V. 2 (1938), pp. 126-149 + tables.
- [3] A. WALD, *Annals of Math. Stat.*, V. 13 (1942), pp. 434-439.
- [4] *The Kelley Statistical Tables*, Macmillan Co., New York, 1938.
- [5] C. M. THOMSON, *Biometrika*, V. 32 (1941-42), pp. 188-9.
- [6] *Ibid.*, pp. 168-181.
- [7] M. MERRINGTON, *Ibid.*, pp. 200-202
- [8] A. M. PEISEL, *Annals of Math. Stat.*, V. 14 (1943), pp. 56-63.
- [9] L. J. COMRIE AND H. O. HARTLEY, *Biometrika*, V. 32 (1941-2), p. 158.
- [10] J. WISHART, *Biometrika*, V. 24 (1932), p. 454
- [11] *Mathematical Tables and Aids to Computation*, V. 1 (1944), pp. 133-5. See also, *ibid.* p. 309.
- [12] *British Association Mathematical Tables*, VI, Part I (1937), pp. 214-263, 272-283.
- [13] E. ANDING, *Sechstellige Tafeln der Besselschen Funktionen imaginären Arguments*. Leipzig, Engelmann, 1911.
- [14] *British Association Mathematical Tables*. For description of these and other tables of Bessel Functions see *Mathematical Tables and Aids to Computation*, V. 1, pp. 224-227.

SOME COMBINATORIAL FORMULAS WITH APPLICATIONS TO PROBABLE VALUES OF A POLYNOMIAL-PRODUCT AND TO DIFFERENCES OF ZERO

By L. C. Hsu

National Southwest Associated University, Kunming, China

1. The main purpose of the paper is to establish some combinatorial formulas concerning the mathematical expectations or probable values of a product of n given polynomials. The problem may be stated more definitely as follows:

Let x_1, \dots, x_n be n non-negative discontinuous variables for which we have assumed that the probability that each x takes a possible value is equally likely, and let $f_1(x), \dots, f_n(x)$ be n given polynomials. Then we shall ask: What is the probable value of the product $f_1(x_1) \cdot \dots \cdot f_n(x_n)$, provided the sum of the variables x_1, \dots, x_n is known? More generally, we may consider the problem with certain restrictions to x such as $a \leq x_i \leq b$, ($i = 1 \dots n$).

By a limiting process¹ it will be found that all the formulas established for the preceding problem can be extended to the case of continuous variables. On this account, it is important to find explicit formulas for the problem merely involving discontinuous variables.

By the definition² of MacMahon, we say that a set of numbers $(x_1 \dots x_n)$ is over all different compositions of m into n parts with each $x \geq k$, if $(x_1 \dots x_n)$ runs over all different integer solutions of the linear equation $x_1 + \dots + x_n = m$ with each $x \geq k$. We shall use the notation $(m, k; x_1 \dots x_n)$, or simply $(m; k; x)$, to denote that a set of numbers $(x_1 \dots x_n)$ is over all different compositions of m into n parts with each $x \geq k$.

The notation $E(m; \delta; [f_1(x)] \dots [f_n(x)])$ will be used to denote the mathematical expectation of the product $f_1(x_1) \dots f_n(x_n)$ in which the sum of n variable quantities x_1, \dots, x_n is known, namely $x_1 + \dots + x_n = m$, and each quantity is a multiple of δ and m is of course a multiple of δ . Thus by the definition³ of mathematical expectations we have

$$(1) \quad E(m; \delta; [f_1] \dots [f_n]) = \left(\sum_{(m/\delta, 1; x)} 1 \right)^{-1} \sum_{(m/\delta, 1, x)} f_1(x_1 \delta) \dots f_n(x_n \delta),$$

where the summation on the right-hand side runs over all different compositions of m/δ into n parts with each $x \geq 1$, and the given constant δ is called a "varying unit", that is the least possible difference between two unequal quantities in $(x_1 \dots x_n)$. If the varying unit approaches zero, $(x_1 \dots x_n)$ will become a set of continuous variables

¹ The limiting process will be illustrated by the proof of corollary 2 of theorem 1 in this paper.

² MacMahon, *Combinatory Analysis*, Vol 1, p 150

³ See for example W. Burnside, *Theory of Probability*, Chap. 4, 13

If $f_1(x) = \dots = f_n(x)$, we may write

$$E(m; \delta; [f]^n) \text{ instead of } E(m; \delta; [f_1] \dots [f_n]).$$

A well-known convention for $\binom{m}{n}$ is also adopted here:

$$\binom{m}{n} = \begin{cases} \frac{m!}{n!(m-n)!}, & \text{if } 0 \leq n \leq m, \\ 0 & \text{otherwise.} \end{cases}$$

2. Lemmas. In order to obtain explicit formulas desired we first establish four lemmas as follows:

LEMMA 1. Let m, r_1, \dots, r_n be non-negative integers. Then

$$(2) \quad \sum_{(m; 0; x)} \prod_{r=1}^n \binom{x_r}{r_r} = \binom{m+n-1}{r_1 + \dots + r_n + n - 1}.$$

PROOF. Construct a generating function:

$$\left(\frac{1}{1-x}\right)^{r_1+1} \dots \left(\frac{1}{1-x}\right)^{r_n+1}, \quad |x| < 1.$$

It is observed that the coefficient of the term $x^{m-(r_1+\dots+r_n)}$ in the expansion of the above product is given by

$$\sum_{(m-r_1-\dots-r_n; 0; x)} \binom{r_1+x_1}{x_1} \dots \binom{r_n+x_n}{x_n} = \sum_{(m; 0; x)} \binom{x_1}{r_1} \dots \binom{x_n}{r_n}.$$

On the other hand we see that the coefficient of the term $x^{m-(r_1+\dots+r_n)}$ in the expansion of $\left(\frac{1}{1-x}\right)^{r_1+\dots+r_n+n}$ is given by

$$\binom{r_1 + \dots + r_n + n + m - (r_1 + \dots + r_n) - 1}{m - (r_1 + \dots + r_n)} = \binom{m+n-1}{2r_1 + n - 1}.$$

Hence the lemma.

LEMMA 2. Let a, b, c, \dots be arbitrary constants, and k_1, k_2, k_3, \dots be positive integers. Then

$$(3) \quad \sum_{(m; 1; x)} \prod_{r=1}^n \left\{ a \binom{x_r}{k_1} + b \binom{x_r}{k_2} + c \binom{x_r}{k_3} + \dots \right\} \\ = n! \sum_{(n; 0; \alpha \beta \gamma \dots)} \binom{m+n-1}{\alpha k_1 + \beta k_2 + \gamma k_3 + \dots + n - 1} \frac{a^\alpha b^\beta c^\gamma}{\alpha! \beta! \gamma!} \dots$$

PROOF. Expanding the left-hand side of (3), we see that the coefficient of the term $a^\alpha b^\beta c^\gamma \dots$ is equal to

$$\frac{n!}{\alpha! \beta! \gamma! \dots} \sum_{(m; 0; x)} \binom{x_1}{k_1} \dots \binom{x_\alpha}{k_1} \binom{x_{\alpha+1}}{k_2} \dots \binom{x_{\alpha+\beta}}{k_2} \binom{x_{\alpha+\beta+1}}{k_3} \dots \binom{x_{\alpha+\beta+\gamma}}{k_3} \dots$$

By lemma 1 it is reduced to

$$\frac{n!}{\alpha! \beta! \gamma! \dots} \binom{m+n-1}{\alpha k_1 + \beta k_2 + \gamma k_3 + \dots + n-1}.$$

Substituting, we get the lemma.

From now on, we shall frequently write $f^{(x)}$ instead of $f(x)$, so that

$$f^{(v)} g^{(0)} + \binom{v}{1} f^{(v-1)} g^{(1)} + \dots + \binom{v}{v} f^{(0)} g^{(v)} = (f+g)^{(v)}.$$

LEMMA 3. Let m, n ($\leq m$) be two positive integers. Then for any given polynomial $f(x)$ of the k th degree we have

$$(4) \quad \sum_{(m;1;x)} f(x_1) \dots f(x_n) = n! \sum_{(n;0,p)} \binom{m+n-1}{S(p)+n-1} \prod_{r=0}^k \frac{[(f-1)^{(v)}]^{p_r}}{p_r!},$$

where

$$f^{(x)} = f(x), \quad S(p) = 1 \cdot p_1 + \dots + k \cdot p_k.$$

PROOF. Since $f(x)$ is a polynomial of the k th degree, there exist $(k+1)$ values β_k, \dots, β_0 such that

$$\beta_k \binom{x}{k} + \dots + \beta_1 \binom{x}{1} + \beta_0 = f(x).$$

By putting $x = 0, 1, \dots, k$, we find successively⁴

$$(5) \quad \beta_r = f^{(v)} - \binom{v}{1} f^{(v-1)} + \dots + (-1)^v \binom{v}{v} f^{(0)} = (f-1)^{(v)},$$

($v = 0, 1, \dots, k$).

The lemma is thus obtained by means of (3).

For convenience, we denote the summation

$$\sum_{(m;1;x)} f_1(x_1) \dots f_n(x_n) \quad \text{by} \quad S(m, [f_1] \dots [f_n]).$$

Thus formula (4) can be rewritten:

$$(4)' \quad S(m, [f]^n) = n! \sum_{(n;0,p)} \binom{m+n-1}{1 \cdot p_1 + \dots + k \cdot p_k + n-1} \frac{\beta_0^{p_0}}{p_0!} \frac{\beta_1^{p_1}}{p_1!} \dots \frac{\beta_k^{p_k}}{p_k!},$$

where $\beta_0, \beta_1, \dots, \beta_k$ are given by (5).

LEMMA 4. Let $f_1(x), \dots, f_n(x)$ be n given polynomials. Then

$$(6) \quad S(m, [f_1] \dots [f_n]) = \frac{1}{n!} \sum_{(v_1, \dots, v_k) \in (1, \dots, n)} (-1)^{n-k} S(m, [f_{v_1} + \dots + f_{v_k}]^n),$$

⁴ Strictly speaking, the relation (5) is established and proved by induction on v , in which a well known combinatorial equality is applied.

where the summation on the right-hand side runs over all different combinations out of $(1 \cdots n)$, ($k = 1 \cdots n$).

For example, if $n = 3$, then all the different combinations out of (123) will be: (1), (2), (3), (12), (13), (23), (123).

PROOF. Consider a typical term

$$\frac{n!}{q_1! \cdots q_t!} S(m, [f_{\nu_1}]^{q_1} \cdots [f_{\nu_t}]^{q_t}),$$

where $1 \leq t \leq n$, $q_1 + \cdots + q_t = n$. Now a necessary and sufficient condition that the term will be contained in the expansion of $S(m, [f_{\nu_1} + \cdots + f_{\nu_k}]^n)$ is $(\nu_1 \cdots \nu_t) \in (\mu_1 \cdots \mu_k)$, i.e. $(\nu_1 \cdots \nu_t)$ is a combination out of $(\mu_1 \cdots \mu_k)$, and for a fixed k , there are $\binom{n-t}{k-t}$ -different combinations of $(\mu_1 \cdots \mu_k)$ satisfying the condition $(\nu_1 \cdots \nu_t) \in (\mu_1 \cdots \mu_k) \in (1 \cdots n)$. Therefore the number of occurrences of the term in the right-hand side of (6) is given by

$$\sum_{r=0}^{n-t} (-1)^r \binom{n-t}{r} = (1-1)^{n-t} = \begin{cases} 0 & \text{if } t < n, \\ 1 & \text{if } t = n. \end{cases}$$

The term vanishes generally except $q_1 = \cdots = q_t = 1$. Hence the right-hand side gives $S(m, [f_1] \cdots [f_n])$.

3. Theorems and corollaries. In the following statement of theorems and corollaries the notation $(x_1 \cdots x_n)$ will be always used to denote a set of undetermined quantities as specified.

THEOREM 1. Suppose that $(x_1 \cdots x_n)$ is a set of natural numbers in which only the sum of the numbers is known, namely $x_1 + \cdots + x_n = m$. Then, for any given polynomial $f(x)$ of the k th degree, the mathematical expectation of $f(x_1) \cdots f(x_n)$ is given by

$$\begin{aligned} & E(m, 1, [f]^n) \\ (7) \quad &= n! \binom{m-1}{n-1}^{-1} \sum_{(n,0;p)} \binom{m+n-1}{S(p)+n-1} \frac{[(f-1)^{(0)}]^{p_0} \cdots [(f-1)^{(k)}]^{p_k}}{p_0! \cdots p_k!}. \end{aligned}$$

PROOF. Let $m' = m + nr$. By lemma 1 we have

$$\sum_{(m;0;x)} \binom{x_1}{0} \cdots \binom{x_n}{0} = \sum_{(m,0;x)} 1 = \sum_{(m',r;x)} 1 = \binom{m' - nr + n - 1}{n - 1}.$$

This is the number of compositions of m' into n parts with each part $\geq r$. In particular, if $r = 1$, we find that the number of compositions of m into n parts is $\binom{m-1}{n-1}$. Thus by (1), the required value is equal to

$$\{S(m, [1]^n)\}^{-1} S(m, [f]^n), \quad \text{i.e.} \quad \binom{m-1}{n-1}^{-1} S(m, [f]^n).$$

The theorem is therefore proved by lemma 3.

COROLLARY 1. Let $(x_1 \cdots x_n)$ be a set of positive quantities, of which the varying unit is δ , and the sum is m . Then, for any given polynomial $f(x)$ of the k th degree, we have

$$E(m, \delta, [f]^n) \\ (8) = n! \left(\left(\frac{m}{\delta} \right) - 1 \right)^{-1} \sum_{(n,0,p)} \left(\left(\frac{m}{\delta} \right) + n - 1 \right) \frac{[(g-1)^{(0)}]^{p_0} \cdots [(g-1)^{(k)}]^{p_k}}{p_0! \cdots p_k!},$$

where

$$g(x) = f(x\delta), \quad S(p) = p_1 + \cdots + kp_k.$$

PROOF. It follows immediately by the relation.

$$E(m, \delta, [f(x)]^n) = E\left(\left(\frac{m}{\delta}\right), 1, [f(\delta x)]^n\right).$$

COROLLARY 2. Let $(x_1 \cdots x_n)$ be a set of non-negatively real numbers, of which the sum is known, namely $x_1 + \cdots + x_n = m$, m being a known real number. Then, for any given polynomial $f(x) = a_0 + \cdots + a_k x^k$, ($a_k \neq 0$) we have

$$E(m, 0, [f]^n) \\ (9) = \frac{(n!)^2}{n} \sum_{(n,0,p)} \frac{m^{1 \cdot q_1 + \cdots + k \cdot q_k}}{(1 \cdot q_1 + \cdots + k \cdot q_k + n - 1)!} \frac{(0! a_0)^{q_0}}{q_0!} \cdots \frac{(k! a_k)^{q_k}}{q_k!}.$$

PROOF. The proof of the corollary depends essentially on the concept that two unequal real numbers may differ by an arbitrary small number h .

Let h be an arbitrary positive number and write $f(xh)/h^k = g(x, h)$, where the number k is the degree of $f(x)$. Then, since

$$\sum_{\nu=0}^n (-1)^\nu \binom{n}{\nu} (n-\nu)^p = \begin{cases} 0 & \text{if } p < n \\ n! & \text{if } p = n \\ \binom{n+1}{2} n! & \text{if } p = n+1, \end{cases}$$

we may write

$$g(\nu, h) = \binom{\nu}{1} g(\nu-1, h) + \cdots + (-1)^\nu \binom{\nu}{\nu} g(0, h) = h^{\nu-k} \{ \nu! a_\nu + h \cdot R_\nu(h) \}, \\ \nu = 0, 1, \cdots, k,$$

where

$$\lim_{h \rightarrow 0} R_\nu(h) = \left\{ \binom{\nu}{0} \nu^{\nu+1} - \binom{\nu}{1} (\nu-1)^{\nu+1} + \cdots + (-1)^{\nu+1} \binom{\nu}{\nu-1} \right\} a_{\nu+1} \\ = \binom{\nu+1}{2} \nu! a_{\nu+1}.$$

Now we pass to the limit $h \rightarrow 0$ in which we assume that h runs through a sequence of real numbers of the form $\frac{m}{N}$, N being a natural number. Thus by corollary 1 we have

$$\begin{aligned} \lim_{h \rightarrow 0} E(m, h, [f]^n) &= n! \lim_{h \rightarrow 0} \sum_{(n; 0; p)} \left(\binom{m}{h} + n - 1 \right) \left(\binom{m}{h} - 1 \right)^{-1} h^{n!} \prod_{r=0}^k \frac{\{\nu! a_r + h \cdot R_r(h)\}^{p_r}}{h^{(k-r)p_r} \cdot p_r!}, \\ &= n! \sum_{(n; 0; p)} \left\{ \lim_{h \rightarrow 0} \frac{\left(\binom{m}{h} + n - 1 \right) (n-1)! \left(\binom{m}{h} - n \right)! h^{s(p)}}{(S(p) + n - 1)! \left(\binom{m}{h} - S(p) \right)! \left(\binom{m}{h} - 1 \right)!} \right\} \\ &\quad \left\{ \lim_{h \rightarrow 0} \prod_{r=0}^k \frac{[\nu! a_r + h \cdot R_r(h)]^{p_r}}{p_r!} \right\}, \\ &= \sum_{(n; 0; p)} \frac{n! (n-1)! m^{s(p)}}{(S(p) + n - 1)!} \prod_{r=0}^k \frac{(\nu! a_r)^{p_r}}{p_r!}, \quad (S(p) = 1 \cdot p_1 + \dots + k \cdot p_k). \end{aligned}$$

Hence the corollary.⁵

COROLLARY 3. Let $(x_1 \dots x_n)$ be a set of positive real numbers under a known condition $a \leq x_1 + \dots + x_n \leq b$, where a ($< b$), b are two positive real numbers. Then, for any given polynomial $f(x) = a_0 + \dots + a_k x^k$, ($a_k \neq 0$), the mathematical expectation of the product $f(x_1) \dots f(x_n)$ which we denote by $E((ab), 0, [f]^n)$ is given by the formula

$$\begin{aligned} (10) \quad E((ab), 0, [f]^n) &= \frac{n! (n-1)!}{b-a} \sum_{(n; 0; q)} \frac{b^{1+S(q)} - a^{1+S(q)}}{(1+S(q)) \cdot (n-1+S(q))!} \\ &\quad \cdot \frac{a_0^{q_0} \dots (k! a_k)^{q_k}}{q_0! \dots q_k!}, \quad (S(q) = q_1 + \dots + k q_k). \end{aligned}$$

PROOF. Since the probability that the sum of x_1, \dots, x_n takes a value between a and b is equally likely, we see that the required mathematical expectation will be the mean of $\int_a^b E(u, 0, [f]^n) du$, that is

$$E((ab), 0, [f]^n) = \frac{1}{b-a} \int_a^b E(u, 0, [f]^n) du.$$

⁵ This corollary can also be proved by means of Dirichlet's integral. In fact, the right-hand side of (9) is given by the quotient of the two integrals:

$$(\int \dots \int f(x_1) \dots f(x_n) dx_1 \dots dx_n) / (\int \dots \int dx_1 \dots dx_n),$$

the integrals being taken over the region: $x_1 + \dots + x_n = m$, $x_1 \geq 0, \dots, x_n \geq 0$.

The formula (10) is therefore obtained by integrating the right-hand side of (9) and dividing it by $(b - a)$.

(On the other hand we find that

$$\lim_{h \rightarrow 0} E((a, a + h), 0, [f]^n) = E(a, 0, [f]^n).$$

Thus shows that the corollary 2 can be also deduced from 3.

THEOREM 2. (A generalization of theorem 1). *Let $f_1(x), \dots, f_n(x)$ be n given polynomials whose degrees do not exceed k . Then we have*

$$(11) \quad E(m, 1, [f_1] \cdots [f_n]) = \sum_{(r_1, \dots, r_t) \in (1, \dots, n)} \sum_{(n, 0; p)} (-1)^{n-t} \frac{\binom{m+n-1}{S(p)+n-1}}{\binom{m-1}{n-1}} \prod_{j=0}^k \frac{[(f_{r_1} \cdots f_{r_t} - 1)^{(j)}]^{p_j}}{p_j!},$$

where

$$f_{r_1} \cdots f_{r_t} = f_{r_1}(x) + \cdots + f_{r_t}(x), \quad S(p) = 1 \cdot p_1 + \cdots + k \cdot p_k.$$

PROOF. In the proof of theorem 1 we have shown that

$$E(m, 1, [f]^n) = \binom{m-1}{n-1}^{-1} S(m, [f]^n).$$

Hence, by similar reasoning and lemma 4 we obtain

$$\begin{aligned} E(m, 1, [f_1] \cdots [f_n]) &= \binom{m-1}{n-1}^{-1} S(m, [f_1] \cdots [f_n]) \\ &= \sum_{(r_1, \dots, r_t) \in (1, \dots, n)} \frac{(-1)^{n-t}}{n! \binom{m-1}{n-1}} S(m, [f_{r_1} + \cdots + f_{r_t}]^n) \\ &= \sum_{(r_1, \dots, r_t) \in (1, \dots, n)} \sum_{(n, 0; p)} (-1)^{n-t} \frac{\binom{m+n-1}{S(p)+n-1}}{\binom{m-1}{n-1}} \prod_{j=0}^k \frac{[(f_{r_1} \cdots f_{r_t} - 1)^{(j)}]^{p_j}}{p_j!}. \end{aligned}$$

Theorem 2 is proved.

COROLLARY 1. *Let δ be a varying unit. Then*

$$(12) \quad E(m, \delta, [f_1] \cdots [f_n]) = \sum_{(r_1, \dots, r_t) \in (1, \dots, n)} \sum_{(n, 0; p)} (-1)^{n-t} \frac{\binom{\left(\frac{m}{\delta}\right) + n - 1}{S(p) + n - 1}}{\binom{m-1}{n-1}^{-1} \prod_{j=0}^k \frac{[(f_{r_1} \cdots f_{r_t} - 1)^{(j)}]^{p_j}}{p_j!}},$$

where

$$g_r(x) = f_r(x\delta), \quad r = 1 \cdots n, \quad \text{and} \quad g_{r_1} = g_{r_1}(x + \frac{1}{2} \cdots + \frac{1}{2} g_{r_l}(x).$$

PROOF. This is almost trivial because of

$$E(m, \delta, [f_1(x)] \cdots [f_n(x)]) = E\left(\binom{m}{\delta}, 1, [f_1(x\delta)] \cdots [f_n(x\delta)]\right).$$

COROLLARY 2.⁶ For any given positive real number m , we have

$$(13) \quad E(m, 0, [x^{p_1}] \cdots [x^{p_n}]) = \frac{p_1! \cdots p_n! (n-1)!}{(p_1 + \cdots + p_n + n - 1)!} m^{p_1 + \cdots + p_n}.$$

PROOF. By a passage to the limit $\delta \rightarrow 0$, we get

$$\begin{aligned} \lim_{\delta \rightarrow 0} E(m, \delta, [f_1] \cdots [f_n]) &= \lim_{\delta \rightarrow 0} \sum_{t=0}^{\infty} \frac{(-1)^{n-t}}{n!} E(m, \delta, [f_{r_1} + \cdots + f_{r_t}]^n) \\ &= \sum_{t=0}^{\infty} \frac{(-1)^{n-t}}{n!} \lim_{\delta \rightarrow 0} E(m, \delta, [f_{r_1} + \cdots + f_{r_t}]^n), \end{aligned}$$

i.e.

$$E(m, 0, [f_1] \cdots [f_n]) = \sum_{t=0}^{\infty} \frac{(-1)^{n-t}}{n!} E(m, 0, [f_{r_1} + \cdots + f_{r_t}]^n).$$

The corollary is then deduced by (9).

THEOREM 3. (Further generalization of theorem 1). Let $(x_1 \cdots x_n)$ be a set of arbitrary integers restricted to the conditions:

$$x_1 + \cdots + x_n = m, \quad a \leq x_i \leq b,$$

where m, a, b are all known integers. Then for any given polynomial $f(x)$, the mathematical expectation of the product $f(x_1) \cdots f(x_n)$ denoted by $E_{(a,b)}(m, 1, [f]^n)$ is given by the following

$$(14) \quad E_{(a,b)}(m, 1, [f]^n) = \frac{\sum_{p=0}^n (-1)^p \binom{n}{p} S(m', [g]^p [h]^{n-p})}{\sum_{p=0}^n (-1)^p \binom{n}{p} \binom{m' - 1}{n - 1}},$$

where

$$g(x) = f(x + b), \quad h(x) = f(x + a - 1), \quad m' = m - (a - 1)n + (a - b - 1)n.$$

PROOF. Define

$$S(m, [f]^n) = 0 \quad \text{for } m < n \quad \text{and} \quad S(m, [f]^0) = \begin{cases} 0 & \text{for } m > 0, \\ 1 & \text{for } m = 0. \end{cases}$$

⁶ This can also be deduced by Dirichlet's integrals.

We shall show that

$$\sum_{\nu=0}^n (-1)^\nu \binom{n}{\nu} S(m', [g]^\nu [h]^{n-\nu}) = \sum_{\substack{(x_1 \dots x_n) \\ a \leq x_i \leq b}} f(x_1) \cdots f(x_n),$$

where on the right-hand side of the expression the set $(x_1 \cdots x_n)$ under the summation runs over all different compositions of m into n parts with each $a \leq x_i \leq b$, ($i = 1 \cdots n$).

We denote the left-hand side of the expression by \mathfrak{S} , then by decomposing $S(m', [g]^\nu [h]^{n-\nu})$ we have

$$\mathfrak{S} = \sum_{\nu=0}^n \sum_{m'=\nu}^{m'} (-1)^\nu \binom{n}{\nu} S(\bar{m}, [f(x+b)]^\nu) S(m' - \bar{m}, [f(x+a-1)]^{n-\nu})$$

Let $f(\bar{x}_1) \cdots f(\bar{x}_n)$ be a term contained in \mathfrak{S} , i.e. $\bar{x}_1 + \cdots + \bar{x}_n = m$, $\bar{x}_1 \geq a, \cdots, \bar{x}_n \geq a$. And we suppose that $\bar{x}_{\nu_i} \geq b+1, \cdots, \bar{x}_{\nu_t} \geq b+1$, where $\nu_i \neq \nu_j$, if $i \neq j$. Then it is found that the number of occurrences of the term in \mathfrak{S} is given by

$$\sum_{t=0}^t (-1)^t \binom{t}{s} = (1-1)^t = \begin{cases} 0 & \text{if } t \geq 1, \\ 1 & \text{if } t = 0. \end{cases}$$

This shows that the term $f(\bar{x}_1) \cdots f(\bar{x}_n)$ of \mathfrak{S} generally vanishes except $a \leq x_i \leq b$. Hence we have

$$\mathfrak{S} = \sum_{\substack{(x_1 \dots x_n) \\ a \leq x_i \leq b}} f(x_1) \cdots f(x_n).$$

Next, we shall find the number of compositions of m into n parts with each $a \leq x_i \leq b$, i.e. the number of terms in \mathfrak{S} . By the result just obtained we see that the number is given by

$$\begin{aligned} & \sum_{\nu=0}^n \sum_{m'=0}^{m'} (-1)^\nu \binom{n}{\nu} \sum_{(m'; 1, x)} 1 \sum_{(m' - \bar{m}, 1, x)} 1 \\ &= \sum_{\nu=0}^n (-1)^\nu \binom{n}{\nu} \left\{ \sum_{m'=0}^{m'} \binom{\bar{m}-1}{\nu-1} \binom{m' - \bar{m} - 1}{n - \nu - 1} \right\} \\ &= \sum_{\nu=0}^n (-1)^\nu \binom{n}{\nu} \binom{m' - 1}{n - 1} \end{aligned}$$

Hence the theorem.

The theorem just proved shows that the mathematical expectation $E(m, 1, [f]^\nu)$ can be expressed by $S(\bar{m}, [g]^\nu)$ and is therefore expressible⁷ in terms of the linear combinations of the coefficients of the polynomial $f(x)$.

COROLLARY 1. Let δ be a varying unit for which $\frac{m}{\delta}, \frac{a}{\delta}, \frac{b}{\delta}$ are all integers. Then

$$E_{(a, b)}(m, \delta, [f(x)]^n) = E_{(\frac{a}{\delta}, \frac{b}{\delta})}\left(\frac{m}{\delta}, 1, [f(\delta x)]^n\right)$$

⁷ See lemmas 3 and 4

COROLLARY 2. Let $f_1(x), \dots, f_n(x)$ be n given polynomials. Then

$$E_{(a,b)}(m, \delta, [f_1] \cdots [f_n]) = \sum_{(r_1, \dots, r_n) \in (1 \cdots n)^n} \frac{(-1)^{n-t}}{n!} E_{(a,b)}(m, \delta, [f_{r_1} + \cdots + f_{r_n}]^n).$$

COROLLARY 3. The number of compositions of m into n parts with each $a_i \leq x_i \leq b_i$, ($i = 1 \cdots n$) is equal to

$$\sum_{r_1=0, \dots, r_n=0}^{1, \dots, 1} (-1)^{r_1 + \dots + r_n} \cdot \binom{m+n-(a_1+\dots+a_n)+(a_1-b_1-1)\nu_1+\dots+(a_n-b_n-1)\nu_n-1}{n-1}$$

PROOF. We have shown that the number of compositions of m into n parts with each $a \leq x \leq b$ is given by

$$\sum_{r=0}^n (-1)^r \binom{n}{r} \binom{m-(a-1)n+(a-b-1)\nu-1}{n-1}.$$

Hence the number of integer solutions of the equation

$$x_{11} + \dots + k_{1n_1} + \dots + x_{s1} + \dots + x_{sn_s} = m$$

with each $a_r \leq x_{r\mu} \leq b_r$ ($\mu = 1 \cdots s$; $r = 1 \cdots n_r$) is given by

$$\begin{aligned} & \sum_{(m; 1, m_1, \dots, m_s)} \sum_{r_1=0}^{n_1} \cdots \sum_{r_s=0}^{n_s} (-1)^{r_1 + \dots + r_s} \prod_{i=1}^s \binom{n_i}{r_i} \\ & \quad \cdot \binom{m_i - (a_i - 1)n_i + (a_i - b_i - 1)\nu_i - 1}{n_i - 1} \\ &= \sum_{r_1=0}^{n_1} \cdots \sum_{r_s=0}^{n_s} (-1)^{r_1 + \dots + r_s} \prod_{i=1}^s \binom{n_i}{r_i} \\ & \quad \cdot \left\{ \sum_{(m; 1, m_1)} \prod_{i=1}^s \binom{m_i - (a_i - 1)n_i + (a_i - b_i - 1)\nu_i - 1}{n_i - 1} \right\} \\ &= \sum_{r_1=0, \dots, r_s=0}^{1, \dots, 1} (-1)^{r_1 + \dots + r_s} \binom{n_1}{r_1} \cdots \binom{n_s}{r_s} \\ & \quad \cdot \binom{m-1 - \sum (a_i - 1)n_i + \sum (a_i - b_i - 1)\nu_i}{n_1 + \dots + n_s - 1}. \end{aligned}$$

The corollary follows at once by putting $n_1 = \dots = n_s = 1$, $s = n$.

The last corollary may be restated as follows: Let there be n stores, b_1, \dots, b_n being the numbers of stocks contained in 1st, 2nd, \dots , n th store respectively. Then m stocks containing at least a_i stocks of the i th store can be taken from these stores in

$$\sum_{r_1=0, \dots, r_n=0}^{1, \dots, 1} (-1)^{r_1 + \dots + r_n} \binom{m+n-1 - \sum a_i + \sum (a_i - b_i - 1)\nu_i}{n-1}$$

— different ways.

We have now established several combinatorial formulas concerning the mathematical expectations of the product $f_1(x_1) \cdots f_n(x_n)$ under certain conditions. Apparently, there are many examples which can be solved by means of the results just obtained. For brevity, we may state a general criterion as follows: The mathematical expectation of a function, $F(x_1, \dots, x_n)$ say, can be estimated by the above mentioned formulas, if and only if 1) the sum of x_1, \dots, x_n is known, and 2) there exist n polynomials $f_1(x), \dots, f_n(x)$ such that $F(x_1, \dots, x_n)$ is proportional to $f_i(x_i)$, ($i = 1 \cdots n$). The undetermined quantities in (x_1, \dots, x_n) may or may not be continuous, if the quantities are discontinuous, the varying unit is necessarily known.

4. Convenient formulas for differences of zero.⁸

Given $f(x) = \beta_0 + \beta_1 x + \cdots + \beta_k x^k$ ($\beta_k \neq 0$) we may write

$$(f - 1)^{(v)} = \sum_{s=0}^k v! \beta_s S_{v,s} = \sum_{s=0}^k \beta_s \Delta^s 0^s,$$

where $S_{v,s}$ is a Stirling number of the second kind, as used by Jordan and defined by

$$v! S_{v,s} = \Delta^s 0^s = \sum_{x=0}^v (-1)^{v-x} \binom{v}{x} x^s,$$

$\Delta^s 0^s$ being in the language of the calculus of finite differences, "a difference of zero".

In terms of the differences of zero, the formulas (7) and (11) may also be restated as follows:

$$(7)' \quad E(m, 1, [f]^n) = \sum_{(n, 0; p)} \frac{(m+n-1)!(m-n)!n!(n-1)!}{(m-S(p))!(S(p)+n-1)!(m-1)!} \\ \times \prod_{v=0}^k \frac{1}{p_v!} (\beta_v \Delta^v 0^v + \cdots + \beta_k \Delta^v 0^k)^{p_v}.$$

$$(11)' \quad E(m, 1, [f_1] \cdots [f_n]) = \sum_{(v_1 \dots v_n) \in (1 \dots n)} (-1)^{n-t} \\ \sum_{(n, 0, p)} \frac{(m+n-1)!(m-n)!(n-1)!}{(m-S(p))!(S(p)+n-1)!(m-1)!} \\ \times \prod_{v=0}^k \frac{1}{p_v!} (B_v \Delta^v 0^v + \cdots + B_k \Delta^v 0^k)^{p_v},$$

where

$$f_v(x) = \beta_{v0} + \cdots + \beta_{vk} x^k, \quad B_\mu = \beta_{1\mu} + \cdots + \beta_{n\mu}, \quad S(p) = 1 \cdot p_1 + \cdots + k p_k.$$

⁸ The methods for obtaining convenient formulas for differences of zero as stated in the first part of this paragraph are similar to those used by Paul S. Dwyer in his paper "The computation of moments with cumulative totals," *Annals of Math Stat*, Vol 9 (1938), pp. 288-303.

The formulas (7)' and (11)' tell us that the difference of zero plays an important rôle in the calculation of mathematical expectations of a polynomial product under known conditions. On account of this fact, we are now going to investigate some recurrence relations and approximations for the differences of zero.

As m is larger than l , we may find a convenient recurrence relation as follows:

$$\begin{aligned} \frac{\Delta^m 0^{m+l}}{m!} &= S_{m,m+l} = \lambda_1(l) \binom{m+l}{l+1} + \lambda_2(l) \binom{m+l}{l+2} + \cdots \\ &\quad + \lambda_l(l) \binom{m+l}{l+l}, \\ (15) \quad \frac{\Delta^m 0^{m+l+1}}{m!} &= S_{m,m+l+1} = \lambda_1(l+1) \binom{m+l+1}{l+1+1} \\ &\quad + \lambda_2(l+1) \binom{m+l+1}{l+1+2} + \cdots + \lambda_{l+1}(l+1) \binom{m+l+1}{2l+2}, \end{aligned}$$

where

$$(l+j)\lambda_{j-1}(l) + j\lambda_j(l) = \lambda_j(l+1), \quad \lambda_0 = \lambda_{l+1}(l) = 0, \quad \lambda_1(l) = \lambda_1(l+1) \equiv 1,$$

and $\lambda_2(l), \dots, \lambda_l(l)$ are all independent of m but depending on l .

Starting with the first equation of (15) and using a well known relation (due to Jordan) $S_{m,n+1} = S_{m-1,n} + m \cdot S_{m,n}$ successively, we get

$$\begin{aligned} S_{m,m+l+1} &= \sum_{\nu=1}^m (m-\nu+1) S_{m-\nu+1, m+l+1-\nu} \\ &= \sum_{j=1}^l \lambda_j(l) \sum_{\nu=1}^m \binom{m+l+1-\nu}{l+j} (m-\nu+1) \\ &= \sum_{j=1}^l \lambda_j(l) \left\{ \sum_{\nu=1}^m \binom{m+l+1-\nu}{l+j+1} (l+j+1) \right. \\ &\quad \left. + \sum_{\nu=1}^m \binom{m+l+1-\nu}{l+j} \cdot j \right\} \\ &= \sum_{j=1}^l \lambda_j(l) \left\{ \binom{m+l+1}{l+j+2} (l+j+1) + \binom{m+l+1}{l+j+1} \cdot j \right\} \\ &= \sum_{j=1}^{l+1} \{ (l+j)\lambda_{j-1}(l) + j\lambda_j(l) \} \binom{m+l+1}{l+j+1}. \end{aligned}$$

The recurrence relation thus follows.

By successive applications of the relation $\lambda_j(l+1) = (l+j) \cdot \lambda_{j-1}(l) + j \cdot \lambda_j(l)$, after n th time say, we may express $\lambda_j(l)$ as a linear combination of $\lambda_j(l-n), \dots, \lambda_{j-n}(l-n)$, but the coefficients are too complicated.

For $t \leq 9$, by applying the recurrence relation as obtained above, the coefficients may be exhibited as follows:

t	$\lambda_2(t)$	$\lambda_3(t)$	$\lambda_4(t)$	$\lambda_5(t)$	$\lambda_6(t)$	$\lambda_7(t)$	$\lambda_8(t)$	$\lambda_9(t)$
1								
2	3							
3	10	15						
4	25	105	105					
5	56	490	1260	945				
6	119	1918	9450	17325	10395			
7	246	6825	56980	190575	270270	135135		
8	501	22935	302995	1636635	4099095	4729725	2027025	
9	1012	74316	1487200	12122110	47507460	94594500	91891800	34459425

For example, when $t = 4$ we have, according to the table:

$$\Delta^n 0^{n+4} = \left\{ \binom{n+4}{5} + 25 \binom{n+4}{6} + 105 \binom{n+4}{7} + 105 \binom{n+4}{8} \right\} m!$$

We shall now proceed to find some approximations for $S_{n,n+t}$ and $\Delta^n 0^{n+t}$. Firstly, we may write

$$S_{n,n+t} = \frac{\Delta^n 0^{n+t}}{n!} = \lambda_1(t) \binom{n+t}{t+1} + \dots + \lambda_t(t) \binom{n+t}{2t}$$

According to the recurrence relation we have

- (i) $\lambda_t(t) = (2t-1)\lambda_{t-1}(t-1)$,
- (ii) $\lambda_{t-1}(t) = 2(t-1) \cdot \lambda_{t-2}(t-1) + (t-1) \cdot \lambda_{t-1}(t-1)$,
- (iii) $\lambda_{t-2}(t) = (2t-3) \cdot \lambda_{t-3}(t-1) + (t-2) \cdot \lambda_{t-2}(t-1)$.

Hence

$$\lambda_t(t) = \frac{(2t)!}{t! \cdot 2^t}, \quad \lambda_{t-1}(t) = (t-1)! 2^{t-1} \sigma(t),$$

$$\lambda_{t-2}(t) = 2^{t-2} \sum_{j=0}^{t-3} (t-2-j)! (t-2-j) \cdot (t-1.5)_j \cdot \sigma(t-1-j),$$

where

$$\sigma(k) = \sum_{x=1}^{k-1} \frac{x}{2^{2x}} \binom{2x}{x}, \quad (t-1.5)_j = (t-1.5)(t-2.5) \cdots (t-j-0.5)$$

Evidently, the orders of $\binom{n+t}{2t-1}$, $\binom{n+t}{2t-2}$, \dots , $\binom{n+t}{t+1}$ are all less than $2t$ as n tends to infinity.

Now, it can be easily found that

$$\begin{aligned}
 \lambda_t(t) \binom{n+t}{2t} &= \frac{(2t)!}{t! 2^t} \binom{n+t}{2t} = \left(\frac{n^2}{2}\right)^t \frac{1}{t!} \left(1 + \frac{t}{n}\right) \cdots \left(1 - \frac{t-1}{n}\right) \\
 &= \frac{1}{t!} \left(\frac{n^2}{2}\right)^t \left(1 + \frac{t}{n}\right) \left(1 - \frac{1}{n^2}\right) \cdots \left(1 - \frac{(t-1)^2}{n^2}\right) \\
 &= \frac{1}{t!} \left(\frac{n^2}{2}\right)^t \left\{1 + \frac{t}{n} - \frac{(t-1)t(2t-1)}{6n^2} + O\left(\frac{1}{n^3}\right)\right\}; \\
 \lambda_{t-1}(t) \binom{n+t}{2t-1} &= \frac{2^t \cdot t!}{n-t+1} \binom{n+t}{2t} \sigma(t) = \frac{2^{2t} t! t!}{(n-t+1) \cdot (2t)!} \sigma(t) \binom{n+t}{2t} \lambda_t(t) \\
 &= \frac{1}{t!} \left(\frac{n^2}{2}\right)^t \left\{ \frac{2^{2t} \sigma(t)}{n \binom{2t}{t}} + \frac{(2t-1) \cdot 2^{2t} \sigma(t)}{n^2 \binom{2t}{t}} + O\left(\frac{1}{n^3}\right) \right\}; \\
 \lambda_{t-2}(t) \binom{n+t}{2t-2} &= \frac{1}{t!} \left(\frac{n^2}{2}\right)^t \left\{ \frac{2^t \cdot t! \cdot \lambda_{t-2}(t)}{n^2 \cdot (2t-2)!} + O\left(\frac{1}{n^3}\right) \right\}.
 \end{aligned}$$

Hence, we may write

$$(16) \quad S_{n,n+t} = \frac{1}{t!} \left(\frac{n^2}{2}\right)^t \left\{1 + \frac{\rho_1}{n} + \frac{\rho_2}{n^2} + O\left(\frac{1}{n^3}\right)\right\},$$

where

$$\begin{aligned}
 \rho_1 &= t + 2^{2t} \binom{2t}{t}^{-1} \sigma(t) = t + \frac{4^t \cdot t! t! \sigma(t)}{(2t)!}, \\
 \rho_2 &= -\frac{1}{6} t(t-1)(2t-1) + \frac{4^t \cdot t! t! (2t-1)}{(2t)!} \sigma(t) + \frac{2^t \cdot t!}{(2t-2)!} \lambda_{t-2}(t).
 \end{aligned}$$

Moreover, it can be shown by Wallis' formula that

$$\sqrt{\frac{x-1}{\pi}} < \frac{x}{2^{2x}} \binom{2x}{x} < \sqrt{\frac{x}{\pi}}, \quad (x = 1, 2, 3, \dots).$$

Thus we have

$$\begin{aligned}
 \sigma(t) &= \sum_{x=1}^{t-1} \frac{x}{2^{2x}} \binom{2x}{x} < \sum_{x=1}^{t-1} \sqrt{\frac{x}{\pi}} < \int_1^t \sqrt{\frac{x}{\pi}} dx = \frac{2}{3\sqrt{\pi}} (t^{\frac{3}{2}} - 1); \\
 \sigma(t) &= \sum_{x=1}^{t-1} \frac{x}{2^{2x}} \binom{2x}{x} > \sum_{x=0}^{t-2} \sqrt{\frac{x}{\pi}} > \int_0^{t-2} \sqrt{\frac{x}{\pi}} dx = \frac{2}{3\sqrt{\pi}} (t-2)^{\frac{3}{2}}.
 \end{aligned}$$

Again, by Wallis' formula we have

$$\sqrt{\pi t} < 4^t \binom{2t}{t}^{-1} < \sqrt{\frac{\pi t^3}{t-1}}.$$

Combining these inequalities we get

$$\frac{2}{3}\sqrt{t} (t-2)^{\frac{1}{2}} < 4^t \binom{2t}{t}^{-1} \sigma(t) < \frac{2}{3}\sqrt{\frac{t^2}{t-1}} (t^{\frac{1}{2}} - 1),$$

where

$$\frac{2}{3}\sqrt{t} (t-2)^{\frac{1}{2}} \sim \frac{2}{3}t^{\frac{1}{2}}, \quad \frac{2}{3}\sqrt{\frac{t^2}{t-1}} (t^{\frac{1}{2}} - 1) \sim \frac{2}{3}t^{\frac{1}{2}}.$$

Therefore,

$$t + \frac{2}{3}\sqrt{t}\sqrt{(t-2)^{\frac{1}{2}}} < \rho_1 < t + \frac{2}{3}\sqrt{t+1}(\sqrt{t^{\frac{1}{2}}} - 1), \quad \rho_1 \sim \frac{2}{3}t^{\frac{1}{2}}.$$

Next, by Stirling's formula

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \left\{1 + \frac{1}{12n} + \frac{1}{288n^2} + o\left(\frac{1}{n^3}\right)\right\}$$

we obtain

$$\Delta^n 0^{n+t} = \left(\frac{n}{e}\right)^n \left(\frac{n^2}{2}\right)^t \frac{\sqrt{2\pi n}}{t!} \left\{1 + \frac{\theta_1}{n} + \frac{\theta_2}{n^2} + o\left(\frac{1}{n^3}\right)\right\} \sim \frac{n^{n+2t}\sqrt{2\pi n}}{e^n \cdot 2^t \cdot t!} \left(1 + \frac{2t^2}{3n}\right),$$

where

$$\theta_1 = \rho_1 + \frac{1}{12} \sim \frac{2}{3}t^{\frac{1}{2}},$$

$$\theta_2 = \rho_2 + \frac{1}{12}\rho_1 + \frac{1}{288}.$$

REFERENCES

- [1] K. MACMAHON, *Combinatory Analysis*. Cambridge Univ. press (1916).
- [2] W. BURNSIDE, *Theory of Probability*, Cambridge (1928)
- [3] PAUL S. DWYER, "Combined expansions of products of symmetric power sums and sums of symmetric power products with application to sampling," *Annals of Math. Stat.*, Vol. 9 (1938), p. 1.
- [4] PAUL S. DWYER, "Computation of moments with cumulative totals," *Annals of Math. Stat.*, Vol. 9 (1938), p. 288.
- [5] JOHN RIORDAN, "Moment recurrence relations for distributions," *Annals of Math. Stat.*, Vol. 8 (1937), p. 137.

RANGES AND MIDRANGES¹

By E. J. GUMMEL

New School for Social Research, New York City

1. Introduction. In the following the generating functions of the extremes are studied in order to determine the nature of the distributions of the ranges and the midranges.

A large sample of size n is considered to be drawn from an unlimited symmetrical continuous distribution with zero mean. The difference and the sum of the largest and of the smallest observation, the extremes, are called *range* and *midrange*. R. A. Fisher and L. H. C. Tippett [2] have established the limiting distributions of the largest and of the smallest member of a sample. The exact conditions under which these distributions hold have been given by R. von Mises [4]. For a normal distribution, L. H. C. Tippett [7] has calculated the numerical values of the mean range and the first four moments of the range for sample sizes varying from 2 to 1000. He has shown that, for sample sizes exceeding 200, the correlation between the largest and the smallest observation may be neglected. Later, E. S. Pearson [5] has calculated the probability function of the range for small samples ($n = 2$ to 20) taken from a normal population. These calculations are very laborious. Recently, W. E. Deming [1] has applied the range to quality control.

The concepts "extremes", "range" and "midrange" allow a simple generalization. Let x_m and $x_{\bar{m}}$ be the m th observation in increasing and in decreasing magnitude, henceforth called m th value "from below" and "from above". As long as the index m is small compared to the sample size n , the m th values under consideration are extremes. The difference and the sum of the m th extreme observations are called the m th *range* and the m th *midrange*. We will investigate the asymptotic distributions of the m th extremes, of the m th range, and of the m th midrange. Assuming that the number of observations is very large, the correlation between the largest and the smallest observation may be neglected. Then the m th range and the m th midrange are the difference and the sum of two independent variates, the m th extremes.

It was found that the distribution of the m th range is skew and the distribution of the m th midrange is of the generalized logistic type, which is symmetrical. For m increasing the distributions of the m th extremes, the m th ranges, and the m th midranges converge toward normality.

2. Generating functions of the m th extremes. Let $\varphi(x)$ be an initial continuous symmetrical distribution with mean zero; let u_m be the most probable m th value from above; let α_m be defined by

$$(1) \quad \alpha_m = \frac{n}{m} \varphi(u_m).$$

¹ Research done with the support of a grant from the American Philosophical Society

Under conditions given in an earlier paper [3] the distributions $f_m(x_m)$ and ${}_mf_{(m)}x$ of the m th extremes are, for large n and small m

$$(2) \quad f_m(x_m) = \alpha_m \frac{m^m}{(m-1)!} e^{-my_m - me^{-y_m}}, \quad {}_mf_{(m)}x = f_m(-x_m)$$

where

$$(3) \quad y_m = \alpha_m(x_m - u_m)$$

is a reduced extreme m th value.

The moment generating function $G_m(t)$ of the m th extreme value from above obtained from the preceding equations is

$$G_m(t) = \frac{m^m}{(m-1)!} e^{u_m t} \int_{-\infty}^{+\infty} e^{-my_m + y_m(t/\alpha_m) - me^{-y_m}} dy_m.$$

Introducing

$$e^{-y_m} = z$$

the integral becomes

$$\int_0^{\infty} z^{m-(t/\alpha_m)-1} e^{-mz} dz = \Gamma\left(m - \frac{t}{\alpha_m}\right) m^{-m+(t/\alpha_m)}$$

whence

$$(4) \quad G_m(t) = e^{u_m t} m^{t/\alpha_m} \Gamma\left(m - \frac{t}{\alpha_m}\right) / \Gamma(m).$$

To obtain the moments of the m th extreme value from above, the fundamental property of the Gamma function is used; whence

$$G_m(t) = e^{(t/\alpha_m)(u_m \alpha_m + l g m)} \prod_{\nu=1}^{m-1} \frac{m - \nu - \frac{t}{\alpha_m}}{m - \nu} \Gamma\left(1 - \frac{t}{\alpha_m}\right).$$

Finally, by reversing the order of multiplication the moment generating function of the m th extreme value from above becomes

$$(5) \quad G_m(t) = e^{(t/\alpha_m)(u_m \alpha_m + l g m)} \prod_{\nu=1}^{m-1} \left(1 - \frac{t}{\nu \alpha_m}\right) \Gamma\left(1 - \frac{t}{\alpha_m}\right)$$

The mean \bar{x}_m of the m th value from above obtained from (5) is

$$(6) \quad \bar{x}_m = u_m + \frac{1}{\alpha_m} \left(l g m - \sum_{\nu=1}^{m-1} \frac{1}{\nu} + c \right)$$

where $c = .57722$ is Euler's constant.

The seminvariant generating function $L_m(t)$ of the m th extreme value from above becomes from (5) and (6)

$$L_m(t) = \frac{t}{\alpha_m} \left(\sum_{\nu=1}^{m-1} \frac{1}{\nu} - c \right) + \sum_{\nu=1}^{m-1} l g \left(1 - \frac{t}{\nu \alpha_m} \right) + l g \Gamma \left(1 - \frac{t}{\alpha_m} \right)$$

and after expansion

$$(7) \quad L_m(t) = \sum_2^{\infty} \frac{t^\nu}{\nu \alpha_m} S_{r,m}$$

where the sums

$$(8) \quad S_{r,m} = \sum_{k=m}^{\infty} \frac{1}{k^r}; \quad r \geq 2$$

are obtained from the sums

$$(8') \quad S_r = \sum_{k=1}^{\infty} \frac{1}{k^r}$$

which are known from the theory of the Gamma function. The numerical values of the seminvariants of the m th extreme value from above $\lambda_{r,m}$, being the coefficients of $t^r/\nu!$ in the expansion (7) may be calculated from a table of the sums $S_{r,m}$ given in an earlier paper [3].

From the generating functions (4) and (7) of the m th extreme value from above the moment generating function ${}_mG(t)$ and the seminvariant generating function ${}_mL(t)$ of the m th extreme value from below are obtained by the symmetric relation (2) as

$$(9) \quad {}_mG(t) = G_m(-t); \quad {}_mL(t) = L_m(-t)$$

and the mean ${}_m\bar{x}$ of the m th extreme value from below is

$$(6') \quad {}_m\bar{x} = -\bar{x}_m.$$

The seminvariants $\lambda_{m,r}$ and ${}_m\lambda_r$ of the m th extreme value from above and from below are linked by

$$(10) \quad \alpha_m \lambda_{m,r} = (\nu - 1)! S_{r,m} = (-1)^r \alpha_{mm} \lambda_r.$$

The standard errors σ_m and ${}_m\sigma$ of the m th extreme values are

$$(11) \quad \alpha_m \sigma_m = \sqrt{S_{2,m}} = \alpha_{mm} \sigma.$$

This procedure for obtaining the moments of the m th extremes from their distribution (2) parallels closely that used by R. A. Fisher and L. C. H. Tippett [2] for obtaining the moments of the largest and smallest value. The special case $m = 1$ of the formulac (4), (9), (6), (6'), (11), and (10) leads to the moment generating functions, the means, the standard errors, and the higher seminvariants of the largest and of the smallest value given by these authors.

The two parameters u_m and α_m which exist in the distribution (2) of the m th extremes may be calculated from the observations by virtue of equations (6) and (11). Thus, the theoretical distributions (2) may be compared to observations even if the initial distribution $\varphi(x)$ is unknown. For increasing m , the distributions of the m th extremes were shown [3] to converge toward normal distributions, their means and standard deviations being given by (6), (6') and 11.

3. Generating functions of the m th range and the m th midrange. To obtain the characteristics for the m th range and the m th mid-range we state first some general properties of the sum and of the difference of two independent variates x and y , with means \bar{x} and \bar{y} and standard deviations σ_x and σ_y . Let the distributions be $\varphi(x)$ and $\psi(y)$, and let the generating functions of the seminvariants $\lambda_{x,r}$ and $\lambda_{y,r}$ be $L_x(t)$ and $L_y(t)$. We write

$$(12) \quad v = x + y; \quad w = x - y$$

from the sum and the difference of the two variates. Then the means \bar{v} and \bar{w} and the variances σ_v^2 and σ_w^2 are

$$(13) \quad \bar{v} = \bar{x} + \bar{y}; \quad \bar{w} = \bar{x} - \bar{y}; \quad \sigma_v^2 = \sigma_x^2 + \sigma_y^2 = \sigma_w^2$$

and the seminvariant generating functions $L_v(t)$ and $L_w(t)$ are

$$(14) \quad L_v(t) = L_x(t) + L_y(t); \quad L_w(t) = L_x(t) + L_y(-t).$$

The negative sign in the second equation (14) is obtained through the same well-known derivation as used for the first equation (14). Therefore, the even seminvariants of the sum are equal to the even seminvariants of the difference, whereas the odd seminvariants of the sum and of the difference are

$$\lambda_{v,2r+1} = \lambda_{x,2r+1} + \lambda_{y,2r+1}; \quad \lambda_{w,2r+1} = \lambda_{x,2r+1} + (-1)^{2r+1} \lambda_{y,2r+1}.$$

If the distributions $\varphi(x)$ and $\psi(y)$ are symmetrical one to another in the sense

$$(15) \quad \psi(y) = \varphi(-x)$$

then the even seminvariants of the two variates x and y coincide in size and sign and the odd seminvariants coincide in size and differ in sign. Under the condition (15), the even seminvariants of the sum v and the even seminvariants of the difference w are twice the even seminvariants of the variates x or y . The odd seminvariants of the sum v are twice the odd seminvariants of the variate x , and the odd seminvariants of the difference w vanish.

We apply these properties to the m th extreme values and write x_m for x and y_m for y . According to (2) the distribution of the m th extreme from above is symmetrical to the distribution of the m th extreme from below in the sense (15).

The m th range w_m and the m th mid-range v_m are defined by

$$(12') \quad w_m = x_m - y_m; \quad v_m = x_m + y_m.$$

The mean \bar{w}_m of the m th range, the mean \bar{v}_m of the m th mid-range and the respective variances $\sigma_{w_m}^2$ and $\sigma_{v_m}^2$ are, from (6), (6') and (11)

$$(13') \quad \bar{w}_m = 2\bar{x}_m; \quad \bar{v}_m = 0; \quad \sigma_{v_m}^2 = 2\sigma_m^2 = 2m\sigma^2 = \frac{2S_{2,m}}{\alpha_m^2} = \sigma_{w_m}^2.$$

The seminvariant generating functions $L_w(t, m)$ of the m th range and $L_v(t, m)$ of the m th mid-range obtained from (7) and (9) are

$$(14') \quad L_w(t, m) = 2 \sum_{r=2}^{\infty} \frac{t^r}{r\alpha_m^r} S_{r,m} = 2L_m(t); \quad L_v(t, m) = 2 \sum_{r=1}^{\infty} \frac{t^{2r} S_{2r,m}}{2r\alpha_m^{2r}}.$$

The seminvariants of the m th range are twice the seminvariants of the m th extreme values from above. The even seminvariants of the m th mid-range and the even seminvariants of the m th range are twice the even seminvariants of the m th extremes. The odd seminvariants of the m th mid-range vanish. Therefore, the distribution of the m th range is skew, and the distribution of the m th midrange is symmetrical. From the convergence of the m th extremes towards normality it follows that the m th range and the m th mid-range tend also, for increasing indices m , toward normality.

The seminvariants of the range and the mid-range are obtained from (13') and (14') by putting $m = 1$ and omitting the index 1. Therefore, the standard errors σ_w and σ_v of the range and of the mid-range are

$$(13'') \quad \alpha\sigma_w = \frac{\pi}{\sqrt{3}} = \sqrt{2} \sigma_n = \alpha\sigma_v$$

where α is given by (1) and σ_n stands for the standard error of the first or the last value. The skewness $\beta_{1,w}$ of the range and the excess $\beta_2 - 3$ for the range and the mid-range

$$(16) \quad \beta_{1,w} = .61928; \quad \beta_{2,w} - 3 = 1.2 = \beta_{2v} - 3$$

are only one half of the corresponding characteristics of the largest value. The distribution of the range is less skew and less concentrated toward the middle than the distribution of the largest value. The moments (13) and (16) of the range and of the midrange may also be obtained directly from Fisher's and Tippett's results [2] when independence of the two extremes is assumed. The numerical values (16) for the limiting distribution of the range differ considerably from the values

$$(16') \quad \beta_{1,w} = .309; \quad \beta_{2,w} - 3 = .54$$

given by Tippett [7] for $n = 1000$ observations. For increasing n the approach of the distribution of the range toward its limiting distribution is very slow. This is reasonable as the approach of the distribution of the largest value toward the limiting distribution is also very slow.

Until now we considered symmetrical initial distributions. The case of an asymmetrical initial distribution may be dealt with briefly. The most probable m th extreme value from below ${}_m u$ differs then from $-u_m$ and

$$(1') \quad {}_m \alpha = \frac{n}{m} \varphi({}_m u)$$

differs from α_m . The distribution of the m th extreme value from below is for large n and small m [3]

$$(2') \quad {}_m f({}_m x) = {}_m \alpha \frac{m^m}{(m-1)!} e^{m {}_m y - m {}_m y^2}$$

where

$$(3') \quad {}_m y = {}_m \alpha ({}_m x - {}_m u)$$

is a reduced m th extreme value from below. The moment generating function ${}_m G(t)$ of the m th extreme value from below becomes

$$(4') \quad {}_m G(t) = e^{m u t} m^{-(t/m\alpha)} \Gamma\left(m + \frac{t}{m\alpha}\right) / \Gamma(m)$$

and the moment generating functions $G_w(t, m)$ and $G_v(t, m)$ of the m th range and of the m th mid-range of an asymmetrical distribution obtained from (4), (4') and in analogy to (14) are

$$(17) \quad \begin{aligned} G_w(t, m) &= e^{t(u_m - m u)} m^{t((1/\alpha_m) + (1/m\alpha))} \Gamma\left(m - \frac{t}{\alpha_m}\right) \Gamma\left(m - \frac{t}{m\alpha}\right) / \Gamma^2(m) \\ G_v(t, m) &= e^{t(u_m + m u)} m^{t((1/\alpha_m) - (1/m\alpha))} \Gamma\left(m - \frac{t}{\alpha_m}\right) \Gamma\left(m + \frac{t}{m\alpha}\right) / \Gamma^2(m) \end{aligned}$$

Thus, the moments of the m th range and the m th mid-range may easily be computed even for an asymmetrical distribution

4. The distribution of the m th midrange. In the following we return to a symmetrical distribution and establish directly the distribution $f(v_m)$ of the m th midrange. Then, the generating function (14') and the convergence toward normality will be verified

From (12') the distribution of the m th midrange is

$$f(v_m) = \int_{-\infty}^{+\infty} f_m(x_m) {}_m f(v_m - x_m) dx_m.$$

Introducing (2), the equation is written

$$f(v_m) = \alpha_m \frac{m^{2m}}{(m-1)!^2} \int_{-\infty}^{+\infty} e^{-m y_m + m \alpha_m (v_m - x_m + u_m) - m e^{-y_m - m e \alpha_m (v_m - x_m + u_m)}} dy_m$$

Using as before

$$e^{-y_m} = z$$

the integral becomes

$$e^{m \alpha_m v_m} \int_0^{\infty} z^{2m-1} e^{-m z (1 + e^{m u_m})} dz = \frac{e^{m \alpha_m v_m} (2m-1)!}{m^{2m} (1 + e^{\alpha_m v_m})^{2m}}.$$

The distribution of the m th midrange is therefore

$$(18) \quad f(v_m) = \alpha_m \frac{(2m-1)!}{(m-1)!^2} \frac{e^{m \alpha_m v_m}}{(1 + e^{\alpha_m v_m})^{2m}}.$$

The distribution (18) will be shown to lead to the seminvariant generating function (14'). The generating function of the m th midrange obtained from (18) is

$$\frac{(2m-1)!}{(m-1)!^2} \int_{-\infty}^{\infty} e^{\alpha_m v_m (m + (t/\alpha_m))} (1 + e^{\alpha_m v_m})^{-2m} d\alpha_m v_m.$$

Introducing

$$1 + e^{\alpha_m v_m} = u^{-1}; \quad du = -u^2 e^{\alpha_m v_m} d\alpha_m v_m$$

the integral is rewritten

$$\begin{aligned} \frac{(2m-1)!}{(m-1)!^2} \int_0^1 (1-u)^{m+(t/\alpha_m)-1} u^{-m-(t/\alpha_m)-1+2m-2} du \\ = \frac{\Gamma(2m)}{\Gamma^2(m)} \frac{\Gamma\left(m + \frac{t}{\alpha_m}\right) \Gamma\left(m - \frac{t}{\alpha_m}\right)}{\Gamma(2m)}. \end{aligned}$$

Consequently, the moment generating function of the m th midrange is

$$(19) \quad G_v(t, m) = \frac{\Gamma\left(m + \frac{t}{\alpha_m}\right) \Gamma\left(m - \frac{t}{\alpha_m}\right)}{\Gamma^2(m)}.$$

This expression is identical with the product of the moment generating functions (4) and (9) of the extreme m th values. Therefore, equation (18) is the distribution, and equation (14') is the seminvariant generating function of the m th midrange.

For $m = 1$ the distribution (18) of the mid-range becomes the so-called logistic distribution which is commonly written

$$(18') \quad f(v) = \frac{\alpha e^{-\alpha v}}{(1 + e^{-\alpha v})^2}.$$

Accordingly, (18) may be called the *generalized logistic distribution*. The probability $F(v)$ of a value equal to, or less than, v obtained from (18') is

$$(20) \quad F(v) = 1/(1 + e^{-\alpha v}).$$

Therefore the distribution $f(v)$ may be expressed by the probability $F(v)$ through

$$(18'') \quad f(v) = \alpha F(v)(1 - F(v)).$$

Formula (20) considered as a growth function plays a rôle in population statistics where it was introduced about 100 years ago by Verhulst [8]. Recently, it has been widely used by R. Pearl [6]. In his treatment, the value v stands for the time, and the function $F(v)$ stands for the relative size of the population at time v compared to its alleged asymptotic size.

In the following, the influence of m on the distribution (18) is studied. The distribution $\phi_m(z)$ of the reduced m th mid-range

$$(21) \quad z = \alpha_m \sqrt{m} v$$

is, from (18)

$$(22) \quad \phi_m(z) = \frac{1}{\sqrt{m}} \frac{(2m-1)!}{(m-1)!^2} \frac{e^{z^2/\sqrt{m}}}{(1 + e^{z^2/\sqrt{m}})^{2m}}.$$

The probability density of the mean m th reduced mid-range increases with m . Indeed,

$$\frac{\phi_{m+1}(0)}{\phi_m(0)} = \frac{2m+1}{2\sqrt{m(m+1)}} > 1.$$

Therefore, the standard error of the reduced m th mid-range decreases with increasing index m . This is reasonable as the m th mid-range is an estimate of the median for the initial distribution. The larger m , the nearer are the m th extreme values to the median, and the better is the estimate.

To verify that the distribution (18) of the m th mid-range tends toward normality for large indices m equation (22) is rewritten

$$\lg \phi_m(z) = \lg \phi_m(0) + z\sqrt{m} - 2m \lg \left(\frac{1 + e^{z^2/\sqrt{m}}}{2} \right).$$

Expansion of the exponential and the logarithm leads, if we neglect the third and higher powers of the deviation z , to

$$\lg \phi_m(z) = \lg \phi_m(0) + z\sqrt{m} - 2m \left(\frac{z^2}{2\sqrt{m}} + \frac{z^2}{4m} - \frac{z^2}{8m} + \dots \right)$$

whence by virtue of (11)

$$f(v) = \text{Const } e^{-\alpha_m^2 m v^2 / 4}.$$

The distribution of the m th mid-range becomes normal for large indices m . The mean is zero, and the standard deviation is

$$(23) \quad \sigma_{v_m} = \frac{\sqrt{2}}{\alpha_m \sqrt{m}}.$$

This is in accordance with the statement (13'), as $S_{2,m}$ tends with increasing m toward $1/m$.

5. Summary. For initial symmetrical distributions the seminvariant generating functions (14') of the m th range and the m th mid-range are obtained from the seminvariant generating functions (7) and (9) of the m th extreme values from above and from below. As the two m th extreme values are supposed to be independent our results hold only for very large, sample sizes. The

even seminvariants of the m th mid-range and of the m th range coincide. The distribution of the m th range is skew and the distribution of the m th mid-range is the generalized symmetrical logistic distribution. For increasing indices m the distributions of the m th extremes, the m th ranges, and the m th mid-ranges converge toward normality.

REFERENCES

- [1] W. F. DIXON, "Some principles of the Shewhart methods of quality control," *Mech. Eng.*, Vol. 66 (1944), pp. 173-177.
- [2] R. A. FISHER AND L. H. C. TIPPETT, "Limiting forms of the frequency distribution of the largest or smallest member of a sample," *Proc. of the Camb. Phil. Soc.*, Vol. 24 (1928), Part 2, pp. 180-190.
- [3] E. J. GUMBEL, "Les valeurs extrêmes des distributions statistiques," *Ann. de l'Institut Henri Poincaré*, Vol. 4 (1935), pp. 115-158.
- [4] R. VON MISES, "La distribution de la plus grande de n valeurs," *Revue Math. de l'Union Interbalkanique*, Vol. 1 (1936), pp. 1-20.
- [5] E. S. PEARSON, "The probability integral of the range in samples of n observations from a normal population," *Biometrika*, Vol. 32 (1942), Parts 3 and 4, pp. 301-308.
- [6] R. PEARL, *Medical Biometry and Statistics*—3rd Edition, W. M. Saunders Co., Philadelphia and London, 1940.
- [7] L. H. C. TIPPETT, "On the extreme individuals and the range of samples taken from a normal population," *Biometrika*, Vol. 17 (1925) Parts 3 and 4, pp. 261-387.
- [8] P. F. VARIANLEY, "Recherches mathématiques sur la loi d'accroissement de la population," *Académie de Bruxelles*, Vol. 18 (1845), pp. 1-38.

NOTES

This section is devoted to brief research and expository articles on methodology and other short items.

NOTE ON ASYMPTOTIC VALUE OF PROBABILITY DISTRIBUTION OF SUM OF RANDOM VARIABLES WHICH ARE GREATER THAN A SET OF ARBITRARILY CHOSEN NUMBERS

BY BRADFORD F. KIMBALL

New York Public Service Commission, New York City

The purpose of this note is to present the following theorem which the author needed in connection with a computational problem. Since the theorem has general implications for statistical theory which do not seem to have been brought out heretofore, readers of this journal may find it of interest.

THEOREM: *In Euclidean space of n dimensions with coordinates x_i ($i = 1, 2, \dots, n$) let a_i be n constants whose sum is u_0 , and consider a hyperplane*

$$(1) \quad x_1 + x_2 + \dots + x_n = u, \quad u \geq u_0.$$

Let k denote a positive constant, and $I(u)$ the $n - 1$ fold integral defined by

$$(2) \quad I(u) = \int \dots \int \exp. \left[-k \left(\sum_{i=1}^n x_i^2 \right) \right] dx_1 dx_2 \dots dx_{n-1}$$

taken over that part of the hyperplane for which $x_i \geq a_i$. Then

$$(3) \quad \lim_{u \rightarrow \infty} \sqrt[n]{n} e^{k u^2 / n} I(u) = (\pi/k)^{(n-1)/2}$$

PROOF: The integral may be reduced to another integral by reduction of the quadratic form

$$\sum_{i=1}^n x_i^2 = \sum_{i=1}^{n-1} x_i^2 + \left(u - \sum_{i=1}^{n-1} x_i \right)^2$$

to a sum of squares y_i^2 such that y_i does not involve x_j for $j < i$. Dropping the superscript $n - 1$ in the notation for Σ and letting the subscript on Σ denote the least value of i involved in the sum, the expansion of this form may be written

$$2 \sum_1 x_1^2 - 2u \sum_1 x_1 + 2 \sum_1 x_1 x_i + u^2, \quad i < j$$

The transformation may be performed progressively as follows.

$$(\sqrt{2} x_1)^2 + 2(x_2 + x_3 + \dots - u)x_1 + [(u - \sum_2 x_i)/\sqrt{2}]^2 = y_1^2$$

giving

$$y_1 = \sqrt{2} x_1 - (u - \sum_2 x_i)/\sqrt{2}$$

Using the remainder of the terms involving x_2 , one can complete the square with terms not involving x_2 , and the value of y_2 is

$$y_2 = x_2 \sqrt{3}/\sqrt{2} - (u - \sum_3 x_i)/\sqrt{6}.$$

Continuing until only terms involving x_{n-1} and u remain,

$$y_{n-1} = x_{n-1} \sqrt{n}/\sqrt{n-1} - u/\sqrt{n(n-1)}.$$

The remaining term in u will be found to be u^2/n .

Making the above transformation, the integral becomes

$$(4) \quad I(u) = [e^{-ku^2/n}/\sqrt{n}] \int \cdots \int \exp. \left[-k \left(\sum_{i=1}^{n-1} y_i^2 \right) \right] dy_1 dy_2 \cdots dy_{n-1}.$$

In order to fix the limits of integration on y , it will be noted that the projection of the critical region of the hyperplane (1) upon the $n-1$ dimensional space in the original variables x_i delineates a region in that $n-1$ space bounded by the $n-1$ hyperplanes

$$x_i = a_i, \quad i = 1, 2, \cdots, n-1$$

and the hyperplane

$$x_1 + x_2 + x_3 + \cdots + x_{n-1} = u - a_n.$$

Hence, if (2) is considered as an iterated integral with the integration performed in the order of the subscripts of x_i , the intervals of integration are

$$\begin{aligned} a_1 &\leq x_1 \leq u - a_n - \sum_2 x_i \\ a_2 &\leq x_2 \leq u - a_n - a_1 - \sum_3 x_i \\ &\dots\dots\dots \\ a_r &\leq x_r \leq u - \sum_{i=1}^r x_i - (u_0 - \sum_{i=r+1}^n a_i) \\ &\dots\dots\dots \\ a_{n-1} &\leq x_{n-1} \leq u - u_0 + a_{n-1}, \end{aligned}$$

where it is recalled that \sum_{r+1} denotes $\sum_{i=r+1}^{n-1}$.

Now transforming to y , and using the general transformation equation

$$(5) \quad y_r = x_r \sqrt{r+1}/\sqrt{r} - (u - \sum_{i=1}^r x_i)/\sqrt{(r+1)r}, \quad 1 \leq r \leq n-1$$

where $\sum_n x_i = 0$ under definition of summation symbol noted above,

$$\text{Lower limit of } y_r = - (u - \sum_{i=1}^r x_i)/\sqrt{(r+1)r} + a_r \sqrt{r+1}/\sqrt{r},$$

$$\text{Upper limit of } y_r = (u - \sum_{i=1}^r x_i) \sqrt{r}/\sqrt{r+1} - (u_0 - \sum_{i=r+1}^n a_i) \sqrt{r+1}/\sqrt{r},$$

since

$$\sqrt{r+1}/\sqrt{r} - 1/\sqrt{(r+1)r} = \sqrt{r}/\sqrt{r+1}.$$

It is not difficult to show that

$$\sum_{r+1} x_i = C_{r+1}(y_i) + (n - 1 - r)u/n$$

where $C_{r+1}(y_i)$ denotes a linear combination of y_i for $i \geq r + 1$, which does not involve the variable u . In other words

$$u - \sum_{r+1} x_i = (r + 1)u/n - C_{r+1}(y_i).$$

Making this substitution, the limits on y_r are found to be

$$\begin{aligned} \text{Lower limit of } y_r &= -(u/n)(\sqrt{r+1}/\sqrt{r}) \\ &\quad + C_{r+1}(y_i)/\sqrt{(r+1)r} + a_r\sqrt{r+1}/\sqrt{r}, \\ \text{Upper limit of } y_r &= (u/n)(\sqrt{r(r+1)}) - C_{r+1}(y_i)\sqrt{r}/\sqrt{r+1} \\ &\quad - (u_0 - \sum_{r+1} a_i)\sqrt{r+1}/\sqrt{r}. \end{aligned}$$

It will now be clear that as u becomes infinite, the limit of the integral in (4), will be the $n - 1$ fold integral taken over the whole of $n - 1$ space. This latter integral is easily evaluated to give $(\pi/k)^{(n-1)/2}$, and the theorem follows.

The following two corollaries, which are restatements of the theorem in less general form, bring out the implications for statistical theory.

COROLLARY 1. With $k = 1/2$ define $F_n(u)$ by

$$(6) \quad F_n(u) = (2\pi)^{-n/2} I(u).$$

The differential $F_n(u)du$ represents then the probability that n random variables x_i taken from a normally distributed population with zero mean and unit standard deviation, fall into a region

$$(7) \quad a_i \leq x_i$$

and have a sum with value in the neighborhood $(\pm \frac{1}{2} du)$ of u .

Recalling that u_0 is the value of u when each x_i has value a_i ,

$$(8) \quad P[a_i \leq x_i] = \int_{u_0}^{\infty} F_n(u) du$$

is the probability that all values of x fall into the region (7). Denoting the normal probability function by $\phi(t)$, corollary 1 implies that

$$(9) \quad \lim_{u \rightarrow \infty} \sqrt{n} F_n(u)/\phi(u/\sqrt{n}) = 1.$$

Since $\phi(u/\sqrt{n})du/\sqrt{n}$ represents the probability distribution of the sum of the n random variables when the condition (7) is removed, certain implications for the theory of statistics emerge. One of these is noted in the example given below. Corollary 1 can be stated in a different form as follows:

COROLLARY 2. If $p_A(u)du$ denotes the probability that the sum of n random varia-

ble from a normally distributed population be in the neighborhood $(\pm \frac{1}{2} du)$ of u , and $p_n(u)$ denotes the probability that the sum of n random variables from the same population that do fall in region (7) have a value in the neighborhood of u , it follows from (9) and the nature of the functions integrated that for arbitrarily small positive δ , a value of u , say u' can be found, sufficiently large, such that for all $u > u'$,

$$(10) \quad (1 - \delta) p_n(u) \leq p_n(u) P[u_1 \leq x_1] \leq p_n(u).$$

Rate of convergence. The author having had occasion to compute $F_n(u)$ for values of n from 2 to 5, and $a_1 = 0$, a table showing the rate of convergence of $F_n(u)$ to its limit for this range of n (and $a_1 = 0$) is shown below. In this table the values of the ratios of the minimum values of $u (= u')$ to the standard deviation of $u (= \sqrt{n})$ are shown for a sequence of values of δ which approaches zero.

Rate of Convergence of $\sqrt{n} F_n(u/\sqrt{n})$ to Unity for $a_1 = 0$

δ	Critical Ratio u'/\sqrt{n}			
	$n = 2$	$n = 3$	$n = 4$	$n = 5$
.5	0.67	1.34	1.92	2.46
.25	1.15	1.95	2.65	3.27
.1	1.64	2.60	3.40	4.11
.05	1.96	3.02	3.89	4.65
.01	2.58	3.85	4.87	5.76
.001	3.29	4.84	6.03	7.08

Example. An example showing the possible bearing of this theorem upon practical considerations of sampling is the following: If in a quality control problem, samples of size 4 were used, and a follow-up of samples which showed a large deviation of sample mean were pursued, a case of a particularly large deviation such as 4 would possibly receive special attention. With sample mean equal to $u/4$, the ratio of this mean to its standard deviation is $u/2$. Turning to the table, in column $n = 4$ it will be noted that the value of δ for $u'/2 = 4$ is somewhat less than .05. Hence from (9) and (10), for $u \geq u'$ ($u' = 8$, $n = 4$),

$$.95 \phi(u/\sqrt{n})/\sqrt{n} < F_n(u) \leq \phi(u'/\sqrt{n})/\sqrt{n}.$$

It follows that

$$.95 \int_{u'}^{\infty} \phi(u/\sqrt{n}) du/\sqrt{n} < \int_{u'}^{\infty} F_n(u) du \leq \int_{u'}^{\infty} \phi(u/\sqrt{n}) du/\sqrt{n}.$$

If one now considers a set of random samples of size 4, the last integral on the right represents the expected proportion of the set which falls into the sub-set A for which $u \geq u'$ (and hence with deviation of mean relative to standard deviation of mean greater than u'/\sqrt{n}). The middle integral represents the expected proportion of the original set which falls into a sub-set B for which $u \geq u'$ and $x_i \geq 0$. It follows from the inequality on the left that the expected proportion of the sub-set A which falls into the sub-set B is greater than 95 per cent. Hence

one infers that the probability is greater than .95 that for a sample showing such a large deviation from the mean ($u/\sqrt{n} = 4, n = 4$) all the constituent elements will have deviations on the same side of the population mean. Thus if all the elements of the sample investigated are found to have deviations on the same side of the population mean, this could *not* be construed as *additional evidence* that the sample indicated an abnormal condition.

This conclusion is weaker than the facts of the example warrant, since it is based upon the *integral* of $F_n(u)$ from u' to infinity. Unfortunately the author does not have data available on the rate of convergence of these integrals.

NOTE ON A MATRIC THEOREM OF A. T. CRAIG

By HAROLD HOTELLING

Columbia University

An extremely elegant theorem given recently by A. T. Craig¹ and applied by him to establish a further theorem on independent χ^2 distributions may be stated as follows:

If A and B are the symmetric matrices of two homogeneous quadratic forms in n variates which are normally and independently distributed with zero means and unit variances, a necessary and sufficient condition for the independence in probability of these two forms is that $AB = 0$.

The proof given that the condition is sufficient is adequate, but Craig's treatment of its necessity consists essentially in its assertion. In view of the growing interest in such quadratic forms, for example in connection with serial correlation, the neatness of this theorem is likely to lead to a wide usefulness. It therefore seems worth while to give a complete proof of the necessity condition.

The form with matrix A is denoted by Q_1 and that with matrix B by Q_2 . The characteristic functions, if defined as $Ee^{i\lambda Q_1}$ and $Ee^{i\mu Q_2}$, are respectively the reciprocals of the square roots of the determinants of the matrices $1 - \lambda A$ and $1 - \mu B$, while the characteristic function for Q_1 and Q_2 together, $Ee^{i(\lambda Q_1 + \mu Q_2)}$, is the reciprocal of the square root of the determinant of $1 - \lambda A - \mu B$. A necessary and sufficient condition for independence is therefore that

$$|1 - \lambda A| \cdot |1 - \mu B| \equiv |1 - \lambda A - \mu B|$$

shall hold identically for all values of λ and μ . Since the determinant of the product of two matrices is the product of their determinants, the left member is the same as

$$|1 - \lambda A - \mu B + \lambda\mu AB|.$$

From this it is immediately obvious that $AB = 0$ implies the independence of the two forms. The converse will now be proved

¹ "Note on the independence of certain quadratic forms," *Annals of Math Stat*, Vol 14 (1943), pp. 195-197

We shall assume therefore that Q_1 and Q_2 are independent, so that the identity holds, and prove that $AB = 0$.

It must not be supposed that Q_1 and Q_2 can by the same linear transformation be reduced to forms in which product terms are absent and only terms in the squares of the variates appear. The available theorems² leading to this canonical form require that at least one of the quadratic forms be non-singular. But it is of the essence of the present situation that both Q_1 and Q_2 be singular, since this is implied by $AB = 0$. It does not appear possible, for example, to reduce to this canonical form the pair $x_1^2 + x_2^2$ and $x_1^2 + 2x_1x_2$.

Nevertheless a real orthogonal transformation can be found reducing Q_1 to

$$d_1x_1^2 + \cdots + d_r x_r^2,$$

where r is the rank of Q_1 . Thus there exists an orthogonal P such that $A = PLP^{-1}$ and $B = PMP^{-1}$, where L and M , when partitioned so as to separate the rows and columns into successive groups of r and $n - r$, are of the forms

$$L = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}, \quad M = \begin{bmatrix} E & C \\ \bar{C}' & F \end{bmatrix}.$$

Here D stands for an r -rowed diagonal matrix having $d_1 \cdots d_r$ in its diagonal, 0 for various matrices whose elements are all zero, and E , C and F for arbitrary matrices of appropriate dimensions. Then

$$|1 - \lambda A| = |P(1 - \lambda L)P^{-1}| = |P| \cdot |1 - \lambda L| \cdot |P^{-1}| = |1 - \lambda L|,$$

and in the same way,

$$|1 - \mu B| = |1 - \mu M|$$

and

$$|1 - \lambda A - \mu B| = |1 - \lambda L - \mu M|.$$

We thus have

$$|1 - \lambda L| \cdot |1 - \mu M| = |1 - \lambda L - \mu M|.$$

From this identity it follows that a pair of forms Q_1^* and Q_2^* , quadratic in a set of variates normally and independently distributed with zero means and unit variances, and having matrices L and M respectively, are independent.

Since $AB = PLMP^{-1}$, the theorem will be proved if we can show that $LM = 0$. Let

$$M_1 = \begin{bmatrix} E & C \\ \bar{C}' & 0 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 0 & 0 \\ 0 & F \end{bmatrix},$$

so that $M = M_1 + M_2$. Since obviously $LM_2 = 0$, we need only to show that $LM_1 = 0$.

² E.g., Bôcher, *Introduction to Higher Algebra*, pp. 169, 305.

Let $Q_2^* = Q' + Q''$, where Q' and Q'' are quadratic forms in the particular normally distributed variates considered, and have matrices M_1 and M_2 respectively. Since Q'' obviously does not involve any of the variates occurring in Q_1^* and since all the variates are independent it follows that Q_1^* is independent of Q'' . Since it has been shown above that Q_1^* is also independent of Q_2^* , it must be independent of the difference $Q_2^* - Q'' = Q'$. Therefore

$$|1 - \lambda L| \cdot |1 - \mu M_1| \equiv |1 - \lambda L - \mu M_1|.$$

We have:

$$|1 - \lambda L| = \prod_{i=1}^r (1 - \lambda d_i).$$

Also,

$$1 - \lambda L - \mu M_1 = \left[\frac{1 - \lambda D - \mu E}{-\mu C'} \middle| \frac{-\mu C}{1} \right].$$

Consequently equating the terms of highest degree in λ on the two sides of the identity

$$\Pi(1 - \lambda d_i) |1 - \mu M_1| \equiv |1 - \lambda L - \mu M_1|$$

yields the identity in μ ,

$$|1 - \mu M_1| \equiv 1,$$

or upon putting $\mu = 1/x$,

$$|M_1 - x| = (-x)^n.$$

Hence all the latent roots of the real symmetric matrix M_1 are zero. Now for a symmetric matrix the sum of the squares of the latent roots equals the sum of the squares of the elements, since both equal the trace of the square of the matrix. Therefore $M_1 = 0$. Consequently $LM_1 = 0$ and the theorem is established.

Since $M_1 = 0$, the following further result is obvious:

Two independent quadratic forms in a set of variates normally and independently distributed with zero means and a common variance can by a transformation be reduced to two forms having no variate in common.

But one of the disjunct sets of variates in the forms as thus reduced is not necessarily independent of the other set. For example, if x_1, x_2, x_3, x_4 are normally distributed with equal variances and any fixed non-vanishing correlation, the same in each of the six pairs, the sets (x_1, x_2) and (x_3, x_4) are not independent of each other, but the forms $(x_1 + x_2)^2$ and $(x_3 - x_4)^2$ are, since $x_1 + x_2$ is uncorrelated with $x_3 - x_4$.

I am indebted to Professor E. R. Lorch in connection with the preparation of the present note.

A NOTE ON THE BEHRENS-FISHER PROBLEM

BY HENRY SCHEFFÉ

Syracuse University

A commonly occurring problem of statistical inference is the comparison of the means of two normal universes when the ratio of their variances is unknown. Let (x_1, \dots, x_m) be a random sample from one normal population with mean α and variance μ , and (y_1, \dots, y_n) , a random sample from another with mean β and variance ν . The problem is then that of making statistical inferences about the difference δ of the means, $\delta = \alpha - \beta$, when the ratio μ/ν is unknown. Convenient tests and confidence intervals are available if one can find a linear form L and a quadratic form Q in the vector $(x_1, \dots, x_m, y_1, \dots, y_n)$ with coefficients independent of the unknown parameters α, β, μ, ν , such that for some positive integer k , the quotient

$$(1) \quad (L - \delta)/(Q/k)^{\frac{1}{2}}$$

has the t -distribution with k degrees of freedom. For this it is sufficient that the following conditions be satisfied for *all* values of the parameters: (i) L and Q are independently distributed, (ii)¹ $E(L) = \delta$, (iii) Q/σ^2 has the χ^2 -distribution with k degrees of freedom, where σ^2 is the variance of L .

In a recent paper [1] the author investigated the Behrens-Fisher problem as delimited by the above three conditions,² and among other results arrived at a simple solution. This solution however does not have the property that the quotient (1) is symmetric, whereby in this note we shall mean the following: A function of the samples and parameters will be called symmetric if it is invariant under permutations of the x 's among themselves and of the y 's among themselves. Let us therefore formulate condition (iv): *the quotient (1) is symmetric*. Since (iv) would be extremely desirable, both for practical and theoretical reasons, and since the author has received several inquiries on this matter, it is considered worth while to outline a proof that conditions (ii) and (iii) imply that (iv) cannot be satisfied, in other words, there exists no "symmetric solution" of the Behrens-Fisher problem within the framework we have imposed. Perhaps this is a simple example of a larger class of problems in which an approach, natural in the light of past developments, forces us to an asymmetric solution.

Suppose (iv) is satisfied. By substituting special values for the vector $(x_1, \dots, x_m, y_1, \dots, y_n)$ and then making permutations allowed by (iv) we find that L and Q must be of the form

$$(2) \quad L = c_1 \sum_i x_i + c_2 \sum_j y_j,$$

$$(3) \quad Q = c_3 \sum_i x_i^2 + c_4 \sum_{i \neq i'} x_i x_{i'} + c_5 \sum_j y_j^2 + c_6 \sum_{j \neq j'} y_j y_{j'} + c_7 \sum_{i,j} x_i y_j,$$

¹ $E(f)$ denotes the expected value of f .

² Although these conditions appear simpler than those in [1] they may be shown equivalent.

where all c 's are independent of parameters, and the range of indices i, i' is from 1 to m , the range of j, j' from 1 to n

Condition (ii) requires that

$$(4) \quad E(L) = \alpha - \beta$$

But from (2),

$$(5) \quad E(L) = c_1 m \alpha + c_2 n \beta.$$

Since (4) and (5) must be satisfied identically in α, β , it follows that $c_1 = 1/m$, $c_2 = -1/n$, and thus the variance of L is

$$(6) \quad \sigma^2 = \mu/m + \nu/n$$

Because of condition (iii) we must have $E(Q/\sigma^2) = k$, and combining this with (6), we have

$$(7) \quad E(Q) = k(\mu/m + \nu/n).$$

However, from (3),

$$(8) \quad E(Q) = c_3 m(\mu + \alpha^2) + c_4(m^2 - m)\alpha^2 + c_5 n(\nu + \beta^2) \\ + c_6(n^2 - n)\beta^2 + c_7 m n \alpha \beta.$$

Equating (7) and (8) gives us an identity in α, β, μ, ν from which we can determine the c 's, and after putting these back in (3) we find that the result may be written

$$(9) \quad Q = k[S_x/(m^2 - m) + S_y/(n^2 - n)],$$

where

$$S_x = \sum_i (x_i - \bar{x})^2, \quad \bar{x} = \sum_i x_i/m, \\ S_y = \sum_j (y_j - \bar{y})^2, \quad \bar{y} = \sum_j y_j/n.$$

The last step of the proof consists of showing that Q defined by (9) violates (ii). Write $u_1 = S_x/\mu$, $u_2 = S_y/\nu$. Then u_1 and u_2 have independent χ^2 -distributions. Now (9) states that $u = Q/\sigma^2$ is of the form $u = a_1 u_1 + a_2 u_2$, where a_1 and a_2 are constants. Let $\phi(t), \phi_1(t), \phi_2(t)$ be the respective characteristic functions of u, u_1, u_2 . Then $\phi(t) = \phi_1(a_1 t) \phi_2(a_2 t)$ because u_1 and u_2 are statistically independent. Since the characteristic function of a χ^2 -variable with r degrees of freedom is $(1 - 2it)^{-r/2}$, it is evident that u has a χ^2 -distribution if and only if $a_1 = a_2 = 1$. From (9),

$$a_1/a_2 = [\mu/(m^2 - m)]/[\nu/(n^2 - n)],$$

and thus a necessary condition (it is also sufficient) for Q/σ^2 to have a χ^2 -distribution is

$$(10) \quad \mu/\nu = (m^2 - m)/(n^2 - n).$$

But (iii) states that Q/σ^2 has a χ^2 -distribution for *all* parameter values. This contradiction completes the proof.

We remark in closing that we have at hand a counter example of practical interest to the statement found in several statistics texts that if z is a normal variable with zero mean and v is an independent unbiased quadratic estimate of the variance of z , then $z/v^{1/2}$ has a t -distribution. The counter example consists of taking $z = \bar{x} - \bar{y} - \delta$ and $v = Q/k$ defined by (9). It may be shown that $z/v^{1/2}$ does not have a t -distribution except in the trivial case (10).

REFERENCE

- [1] H. SCHEFFÉ, *Annals of Math. Stat.*, Vol. 15 (1943), pp. 35-44.

ON MULTIPLE MATCHING WITH ONE VARIABLE DECK

T. N. E. GREVILLE

Bureau of the Census

The problem of card matching has been considered by a number of writers. A complete bibliography has been given by Battin [1], who also published the most general treatment of the subject to date, dealing with the simultaneous matching of any number of decks of arbitrary composition. He considers, however, only the case in which the order of every deck is variable, all possible permutations being equally likely. Some interest attaches to the case in which all the decks but one have fixed orders in relation to one another, especially in connection with radio experiments in telepathy, where a large number of subjects simultaneously attempt to call the same target.

The simplest case is that in which the target for each trial is chosen at random, independently of the other trials. If the target is to be selected from s possibilities, and if p_i denotes the probability that the i th possibility will be chosen as the target, while m_i denotes the number of subjects who call the i th possibility, then the mean value of h , the number of correct calls is, of course,

$$(1) \quad M_h = \sum_{i=1}^s p_i m_i,$$

and the variance is

$$(2) \quad V_h = \sum_{i=1}^s p_i m_i^2 - M_h^2.$$

Evidently, the mean number of hits for a succession of trials is the sum of the means for the individual trials, and the variance is the sum of the variances.

A slightly more difficult problem is presented when the target series is a true "deck": that is, when its composition is determined in advance, only the order being left to chance. Let n denote the number of trials and n_i ($i = 1, 2, \dots, s$)

the number of targets of the i th kind so that $\sum_{i=1}^s n_i = n$. Also let m_{ij} ($j = 1, 2, \dots, n$) denote the number of subjects who call the i th possibility on the j th trial, and let m_i denote the total number of such calls in all the n trials, so that $m_i = \sum_{j=1}^n m_{ij}$. Then, using the ingenious type of counting function introduced by Wilks [6] and improved by Battin [1], we define

$$\Phi = \frac{1}{N} \prod_{j=1}^n \left(\sum_{i=1}^s x_i e^{m_{ij}\theta} \right),$$

where $N = n! / \prod_{i=1}^s n_i!$. We also define an operator K_n , such that if $u = u(x_1, x_2, \dots, x_s)$, then $K_n u$ is the coefficient of $x_1^{n_1} x_2^{n_2} \dots x_s^{n_s}$ in u . Then, if h denotes the number of hits

$$P(h = r) = \text{coefficient of } e^{r\theta} \text{ in } K_n \Phi.$$

Also,

$$E(h^p) = K_n \left. \frac{\partial^p \Phi}{\partial \theta^p} \right|_{\theta=0}.$$

It follows immediately that

$$(3) \quad M_h = \frac{1}{n} \sum_{i=1}^s n_i m_i = \sum_{i=1}^s p_i m_i,$$

where p_i is written for n_i/n . Similarly, it can be shown that

$$E(h^2) = \frac{1}{n} \sum_{i=1}^s \left(n_i \sum_{j=1}^n m_{ij}^2 \right) + \frac{1}{n(n-1)} \sum_{i=1}^s \left[n_i(n_i-1) \sum_{\substack{j,k=1 \\ j \neq k}}^n m_{ij} m_{ik} \right] \\ + \frac{1}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^s \left[n_i n_j \sum_{\substack{k,l=1 \\ k \neq l}}^n m_{ik} m_{jl} \right].$$

If we write $\lambda_{ij} = \sum_{k=1}^n m_{ik} m_{jk}$, it is found that

$$(4) \quad V_h = \frac{1}{n-1} \left[\sum_{i=1}^s p_i (n\lambda_{ii} - m_i^2) - \sum_{i,j=1}^s p_i p_j (n\lambda_{ij} - m_i m_j) \right].$$

It should be noted that it is not necessary that the total number of subjects be the same for each trial.

Certain special cases are of interest. When there is just one subject for each target, as in the ordinary matching of two decks, $\lambda_{ij} = m_i$ for $i = j$, and 0 otherwise; and $\sum_i m_i = n$. Then, if π_i denotes m_i/n and $\rho_i = 1 - \pi_i$, and if $q_i = 1 - p_i$,

$$(5) \quad M_h = n \sum_i p_i \pi_i \quad \text{and} \quad V_h = \frac{n^2}{n-1} [\sum_i p_i q_i \pi_i \rho_i + \sum_{i \neq j} p_i p_j \pi_i \pi_j],$$

a compact expression which is equivalent to the forms previously given by Stevens [1, 5], Greenwood [2], Battin [1], and the author [3].

On the other hand, if the different kinds of targets occur with equal frequency, so that every $p_i = 1/s$ and every $q_i = (s-1)/s$; and if $v_i = \Sigma_k m_{ik}$ and $v = \Sigma_i m_i = \Sigma_k m_k$, the expression (4) becomes

$$(6) \quad Y_h = \frac{n}{s^2(n-1)} [v^2 - n\Sigma_i v_i^2 + s\Sigma_i (n\lambda_{ii} - m_i^2)].$$

REFERENCES

- [1] I. L. BATTIN, "On the problem of multiple matching," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 291-305.
- [2] J. A. GREENWOOD, "The first four moments of a general matching problem," *Annals of Eugenics*, Vol. 10 (1940), pp. 290-292.
- [3] T. N. E. GREVILLE, "The frequency distribution of a general matching problem," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 353-354.
- [4] W. L. STEVENS, "Distribution of entries in a contingency table," *Annals of Eugenics*, Vol. 8 (1938), pp. 238-244.
- [5] W. L. STEVENS, "Tests of significance for extra sensory perception data," *Psychol. Review*, Vol. 46 (1939), pp. 142-150.
- [6] S. S. WILKS, "Statistical aspects of experiments in telepathy," a lecture delivered to the Galois Institute of Mathematics, Long Island University, Dec. 4, 1937.

BOOK REVIEW

"Student's" Collected Papers. Edited by E. S. PEARSON and JOHN WISHART.
Biometrika Office, London. pp. xiv + 224. 1942. 15 shillings.

Since the question is still asked from time to time, it may be well to state that 'Student' was the pen name of William Sealy Gosset. After taking his degree in mathematics and natural sciences at Oxford, Student worked as a brewer for the well-known firm of Messrs. Guinness—in Ireland from 1899 till 1935 and thereafter in London as head of the new Guinness brewery established there. He died in 1937 at the age of 61. Between 1907 and 1937 he published twenty-one papers and a few notes; these are issued in the present volume as a tribute from a group of his relatives and friends.

Student's name is almost universally attached to a single discovery, the t -test, which requires for a complete proof more powerful mathematical methods than he devised. These facts may cause his papers to be regarded as museum pieces by people who have not read them. Actually, in many respects no better model than Student could be suggested for a young statistician today. Since we sometimes speak derisively of courses of lectures in statistics where the methods and ideas are "20 years out of date", it is interesting to find that many of the fundamental ideas considered most in need of emphasis by forward-looking teachers today were in fact emphasized by Student in his writings over 30 years ago.

For example, his classical paper on the t -test, published in 1908, opens as follows: "Any experiment may be regarded as forming an individual of a 'population' of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population."

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong."

The idea is elaborated in a second paper published in the same year. "Note that the indefinitely large population need not actually exist. In Mr. Hooker's case his sample was 21 years of farming under modern conditions in England, and included all the years about which information was obtainable. Probably it could not actually have been made much larger without loss of homogeneity, due to the mixing with farming under conditions not modern, but one can imagine the population indefinitely increased and the 21 years to be a sample from this." We note here a further observation, not always appreciated, that in some lines of research samples which are both large and *homogeneous* do not exist. A clear conception of the relation between sample and population is evident in all his work: one further quotation, the summary from his 1926 paper, *Mathematics and Agronomy*, will suffice. "To sum up, in planning agronomic experiments use plenty of replications and make quite sure your results are capable of being considered to be a random sample of the population about which you wish to draw conclusions."

Another noticeable feature of Student's work is his concern about the extent to which the mathematical model underlying his statistical techniques really represents the facts of the data. Further, he is always ready to give his opinion, from considered thought, as to whether the discrepancies between the model and the facts will affect the conclusions materially. In his first paper, *On the error of counting with a haemocytometer*, he shows that the Poisson distribution applies to the total number of cells or corpuscles which are found in a division of the haemocytometer field. He is careful, however, to point out three ways in which deviations from the Poisson law may occur in practice: (i) there may be interference between the particles, though this is considered unlikely in high dilutions (ii) there may be clumping and (iii) (probably the most important, in his view) the drops taken out for counting may not represent the bulk of the liquid, giving rise to an additional sampling error superimposed on the Poisson error. In a later paper *An explanation of deviations from Poisson's law in practice* (1919), he makes a critical examination of the assumptions necessary for Poisson's distribution, and shows the type of distribution to be expected from the invalidity of one or another of the assumptions.

Similarly, in the opening paragraphs of the 1908 paper on the *t*-test, he notes that the assumption of normality in the parent distribution may not hold in practice. Then follows his opinion on the importance of this assumption: "It appears probable that the deviation from normality must be very extreme to lead to serious error."

His earliest papers deal with attempts to find several exact small-sample distributions for which he and other workers in Britain felt a need. Lacking the mastery of probability theory necessary for obtaining a rigorous solution, he used a combination of mathematics, intelligent guesswork and inferences from repeated samplings which he drew. For example, the steps in his development of the *t*-distribution were as follows:

(i) The distribution of the mean square, s^2 , being unknown to British statisticians at that time, had first to be found. Student calculated the first four moments by algebraic methods. Noticing that the condition for fitting a Pearson Type III curve ($2\beta_2 - 3\beta_1 - 6 = 0$) held exactly, he inferred correctly that s^2 followed a type III curve.

(ii) He was aware that the joint distribution of s^2 and the sample mean \bar{x} was necessary for a solution of the problem and further that the form of the joint distribution depended on the correlation between \bar{x} and s^2 . Having shown by algebra that s^2 is uncorrelated with either \bar{x} or \bar{x}^2 , he assumed (without further comment) that the distributions of \bar{x} and s^2 were independent.

(iii) The final step to the *t*-distribution was performed rigorously.

(iv) The result was then checked against empirical data. He used the heights of 3000 criminals, which provided 750 samples of 4. Another set of 750 samples was supplied by the lengths of the left middle fingers of the same criminals.

In 1908 Student also published the distribution of the sample correlation coefficient r , in a bivariate normal population in which ρ is zero. His interest in

the problem arose from a meeting of the Royal Statistical Society, in which an inquiry was raised about the significance of correlation coefficients derived from small numbers of cases. The authorities present, Yule, Hooker and Edgeworth, ventured their opinion on a specific instance, but had no general answer. Accordingly, Student set out to find the general distribution of r , without assuming ρ zero. His mathematics did not take him far; providing the solution only for samples of 2, where the frequency is zero except at -1 and $+1$. In order to learn what empirical results could teach about the form of the distribution, he then drew four sets of 750 samples, the sets having $\rho = 0, n = 4; \rho = 0, n = 8; \rho = .66, n = 4$, and $\rho = .66, n = 8$, respectively.

He considered first the set having $\rho = 0, n = 4$. Now, for samples of 4, the exact distribution, $C(1 - r^2)^{\frac{1}{2}(n-4)}$, reduces to a rectangular distribution. For his purpose, it would seem that Student had been unlucky in his choice of $n = 4$, since the rectangular distribution appears to furnish little or no clue as to the general form of distribution when n differs from 4. Actually, things could not have turned out better. Referring, as in the case of the distribution of s^2 , to Pearson's curves, he selected a type II on account of the limited range of the distribution of r . The fitted curve was found to be $C\left(1 - \frac{r^2}{1.076}\right)^{0.272}$. This suggested to Student that the true form ought to be $C(1 - r^2)^0$, giving a rectangular distribution, the general form was guessed as $C(1 - r^2)^{\frac{1}{2}(n-4)}$. The samples of 8 were then used merely as corroborative material. One wonders how Student would have fared, both in this case and in the case of s^2 , if the Pearson system of curves had not been familiar to him. The method did not of course provide the much more complex distribution when ρ is not zero; however, Student was able to write down several important properties of the distribution for the guidance of mathematical statisticians who might be interested in the problem.

By similar methods he attempted (1921) to find the frequency distribution and the efficiency of Spearman's rank correlation coefficient. Much later (1936), he remarked in connection with this coefficient that, though it was not fully efficient, "it was so simple that when playing with *other* people's figures, for instance on a railway journey, it was the obvious one to use."

His paper, *The elimination of spurious correlation due to position in time and space*, though brief and incomplete in its treatment of the problem, established him as one of the pioneers in the use of the variate-difference method. His suggestion was to correlate the n th differences of x with those of y for $n = 1, 2, \dots$ until a point was reached where the correlation ceased to change. This limiting correlation would be free from spurious effects due to trend factors. Such methods, he remarks, help to show "whether there really is a close connexion between the female cancer death rate and the quantity of imported apples consumed per head."

His association with field experiments conducted in Ireland for the purpose of developing high yielding varieties of barley suitable for brewing led to an interest in the design of experiments, to which he devoted seven papers. His philosophy

on this subject, though it appeared to be increasingly out of harmony with the philosophy underlying Fisher's methods, was a natural outgrowth of Student's general attitude towards the application of statistical techniques. He believed that, from general experience of soil fertility patterns plus local knowledge of the history and topography of the particular field in question, one could construct a systematic design which would be more accurate than the most appropriate randomized design. Secondly, although systematic designs rely, in nature, as it were, to supply the random element necessary for the application of probability theory, he believed, again from knowledge of field conditions, that the discrepancies between the mathematical model and the true fertility pattern would not seriously vitiate the estimate of error. How far he would have carried the advocacy of systematic arrangements is not clear. Nearly all of his discussions refer to arrangements for comparing only two varieties. Moreover, he stressed repeatedly the necessity for carrying out such trials over several seasons and at a considerable number of places, in order to sample weather and soil variations. In such cases the statistical significance of a result at a single place did not interest Student greatly.

A few quotations of his opinions on more detailed questions illustrate his understanding of the attitude of the farmer and agriculturist and his direct mode of expression. Speaking of experimentation with large plots, he says "it has the advantage that the farmer, who always has a healthy contempt for gardening, may pay some attention to the results." On adjustments to yield data to allow for variations in soil fertility, he writes: "There is a great disadvantage in correcting any figures for position, measurement, etc., errors of cooking, and besides the corrected figures do not represent anything real." Long experience had made him more and more convinced that a serial correlation is inevitable in routine chemical analyses; accordingly, he advises: "The chemist who wishes to impress his clients will therefore arrange to do repetition analyses as nearly as possible at the same time, but if he wishes to diminish his real error he will separate them by as wide an interval of time as possible."

Re-reading of Student's papers has been a keen pleasure. It is hoped that the volume will enjoy a wide circulation amongst statisticians.

W. G. COCHRAN
